Tests of independence by bootstrap and permutation : an asymptotic and non-asymptotic study. Application to neurosciences.

Melisande Albert

Université Nice Sophia Antipolis - UFR Sciences
École Doctorale de Sciences Fondamentales et Appliquées

Thèse

pour obtenir le titre de

Docteur en Sciences

de l'Université Nice Sophia Antipolis

Discipline : Mathématiques

présentée et soutenue par

Mélisande ALBERT

Tests d'indépendance par bootstrap et permutation : étude asymptotique et non-asymptotique. Application en Neurosciences

Tests of independence by bootstrap and permutation: an asymptotic and non-asymptotic study. Application to Neurosciences.

Thèse dirigée par **Patricia REYNAUD-BOURET** et **Magalie FROMONT**

soutenue le 16 novembre 2015

Jury :

M.	Sylvain ARLOT	*Examinateur*	Professeur des Universités
M.	Patrice BERTAIL	*Examinateur*	Professeur des Universités
M.	Gilles BLANCHARD	*Rapporteur*	Professeur des Universités
M.	Jean-François COEURJOLLY	*Examinateur*	Maître de Conférences
Mme.	Magalie FROMONT	*Co-directrice de thèse*	Maître de Conférences
M.	Robert E. KASS	*Rapporteur*	Professeur des Universités
M.	Oleg LEPSKI	*Examinateur*	Professeur des Universités
Mme.	Patricia REYNAUD-BOURET	*Directrice de thèse*	Directrice de Recherche CNRS

Cette thèse a été financée par une bourse doctorale de la région Provence-Alpes-Côte d'Azur, parrainée par le département OBELIA de la filiale AXONIC (du groupe MXM), et a été préparée au Laboratoire Jean Alexandre Dieudonné, Université Nice Sophia Antipolis.

Résumé. Initialement motivée par la détection de synchronisations dans l'analyse des suites de potentiels d'action en neurosciences, cette thèse vise à construire de nouveaux tests d'indépendance non-paramétriques, applicables aux processus ponctuels, ayant de bonnes propriétés à la fois asymptotiques et non-asymptotiques.
Dans un premier temps, nous construisons de tels tests basés sur des approches par bootstrap ou par permutation. Nous étudions leurs performances asymptotiques dans un cadre de processus ponctuels, à travers l'étude du comportement asymptotique des lois conditionnelles des statistiques de test bootstrappée et permutée, sous l'hypothèse nulle ainsi que sous n'importe quelle alternative. Une étude par simulation nous permet de vérifier que ces tests sont bien applicables en pratique, et de les comparer à d'autres méthodes classiques en neurosciences. Ensuite, nous nous intéressons plus particulièrement aux tests par permutation, connus pour leurs bonnes propriétés en termes de niveau non-asymptotique. Les p-valeurs correspondantes, basées sur la notion de coïncidences avec délai, sont implémentées dans une procédure de tests multiples de type Benjamini-Hochberg, appelée méthode *Permutation Unitary Events*, afin de détecter les synchronisations entre suites de potentiels d'action en neurosciences. Nous avons vérifié la validité d'un point de vue pratique de la méthode avant de l'appliquer à de vraies données.
Dans un second temps, nous avons étudié les performances non-asymptotiques des tests par permutation en termes de vitesse de séparation uniforme. Nous construisons une nouvelle procédure de tests agrégés, basée sur l'approche par permutation et une méthode de seuillage par ondelettes dans un cadre de variables aléatoires à densité. Habituellement, le contrôle précis des quantiles nécessite des inégalités de concentration. En nous basant sur les inégalités fondamentales pour les permutations aléatoires de Talagrand, nous démontrons une nouvelle inégalité de concentration de type Bernstein pour des sommes permutées aléatoirement. Cela nous permet alors de majorer la vitesse de séparation uniforme de notre procédure agrégée sur des classes particulières de fonctions, à savoir les espaces de Besov faibles, par rapport à la distance quadratique, et de déduire, au vue de la littérature, que cette procédure semble être optimale et adaptative d'un point de vue minimax.

Abstract. Initially motivated by synchrony detection in spike train analysis in neuroscience, the purpose of this thesis is to construct new non-parametric tests of independence adapted to point processes, with both asymptotic and non-asymptotic good performances.
On the one hand, we construct such tests based on bootstrap and permutation approaches. Their asymptotic performance are studied in a point process framework through the analysis of the asymptotic behaviors of the conditional distributions of both bootstrapped and permuted test statistics, under the null hypothesis as well as under any alternative. A simulation study is performed verifying the usability of these tests in practice, and comparing them to existing classical methods in neuroscience. We then focus on the permutation tests, well known for their good properties in terms of non-asymptotic level. Their p-values, based on the delayed coincidence count, are implemented in a Benjamini-Hochberg type multiple testing procedure, called *Permutation Unitary Events method*, to detect the synchronization occurrences between two spike trains in neuroscience. The practical validity of the method is verified on a simulation study before being applied on real data.
On the other hand, the non-asymptotic performances of the permutation tests are studied in terms of uniform separation rates. A new aggregated independence testing procedure based on the permutation approach, and a wavelet thresholding method is developed in the density framework. Classically, concentration inequalities are necessary to sharply control the quantiles. Based on Talagrand's fundamental inequalities for random permutations, we provide a new Bernstein-type concentration inequality for randomly permuted sums. In particular, it allows us to upper bound the uniform separation rate of the aggregated procedure over particular classes of functions, namely weak Besov spaces, with respect to the quadratic metric and deduce that, in view of the literature, this procedure seems to be optimal and adaptive in the minimax sense.

Key words: Independence test, bootstrap, permutation, randomization, U-statistics, point processes, neuroscience, spike train analysis, synchronization, Unitary Events, trial-shuffling, multiple testing, concentration inequalities, uniform separation rates, adaptive tests, wavelets, weak Besov bodies, aggregated tests.

Remerciements

Me voilà donc arrivée au temps des remerciements, qui marquent la fin de trois années passionnantes, pleines de travail, de rencontres enrichissantes aussi bien sur le plan personnel que professionnel, de découvertes et de voyages. Je tiens à remercier tous ceux qui ont contribué, de près ou de loin, à cette thèse.

Avant tout, je tiens à remercier très sincèrement mes deux directrices de thèse, Patricia et Magalie, sans qui je n'en serais pas là aujourd'hui. Vous m'avez fait découvrir le monde de la recherche, et m'y avez vraiment donné goût. Merci infiniment pour votre enthousiasme permanent et votre patience, votre soutien dans les moments de doutes, vos précieux conseils et vos relectures (ce manuscrit de thèse vous doit beaucoup), votre complicité (c'est une grande chance d'avoir des directeurs solidaires), et votre disponibilité tout au long de ces trois années. Bref, merci d'avoir été des encadrantes si exceptionnelles !

Je tiens également à remercier mes rapporteurs. Robert E. Kass, thank you so much for accepting to review my PhD thesis, and for your interest in my work; it is a great honor to count you among my thesis committee. Gilles Blanchard, merci également d'avoir accepté de rapporter ma thèse, je suis vraiment très honorée (et émue) de l'intérêt que vous y avez porté. Mes remerciements s'adressent également aux membres de mon jury de thèse : Sylvain Arlot, Patrice Bertail, Jean-François Coeurjolly et Oleg Lepski, merci d'avoir accepté de faire partie de ma soutenance.

Je souhaite témoigner ma reconnaissance envers tous ceux qui m'ont aidée au départ de cette thèse ; Jean-Louis Divoux pour son soutien et le parrainage de ma thèse par Obelia (de la filiale Axonic du groupe MXM), ainsi que Hélène Politano pour son aide précieuse.

Cette thèse a bénéficié de nombreuses discussions fructueuses. En premier lieu, je tiens particulièrement à remercier Yann Bouret, d'avoir porté de l'intérêt à ces travaux, et accepté d'en faire partie. Merci également à Franck Grammont pour sa pédagogie et ses nombreuses explications sur la problématique en neurosciences. Je remercie également Bernard Delyon pour nos discussions autour de la concentration, et Vincent Rivoirard pour ses conseils avisés en terme de théorie minimax.
J'ai également une petite pensée pour certains de mes anciens professeurs qui m'ont donné le goût des mathématiques, dont notamment, Mme Mondini, Mme Petri et Mlle Fiole, et plus récemment, Magalie Ribot, Florent Berthelin, Erwan Aubry et David Chiron.

Je dois dire également que le Laboratoire Jean Alexandre Dieudonné a été un formidable laboratoire d'accueil. Tout d'abord, un grand merci à Christine, Thomas, François, Yannick

et tous les membres de l'équipe Probabilité et Statistiques de m'avoir accueillie et aidée à de nombreuses reprises. Et un grand merci aussi à tous les autres membres du LJAD, rencontrés autour d'un café à 10h, d'un délicieux gâteau fait maison lors du GDR goûter ou d'un mot croisé à midi. Un grand merci à Francine pour s'être si bien occupée de notre équipe pendant ces années, ainsi que toute l'équipe administrative du LJAD. Merci également à mon voisin de bureau, Philippe Maisonobe, toujours là pour égayer ma journée.
Je tiens à remercier également tous les Rennais que j'ai eu l'occasion de rencontrer à l'ENSAI, et à Rennes 2, pour leur accueil chaleureux à chaque fois où je suis montée en Bretagne.

Je n'ai pas suffisamment de place pour remercier tous les doctorants, du LJAD et d'ailleurs, anciens et nouveaux ! Que de bons moments passés ensembles, au GDR GOT (et art cinématographique en tous genres), pendant les soirées d'été sur la plages (à jouer au "Nathalie stone game"), pendant les journées d'hiver au ski (à réviser sur les télésièges), aux colloques des doctorants... J'ai une pensée spéciale pour un grand amateur de kawa du LPMC. Vous avez rendu ces années de thèse inoubliables. Merci !
J'ai également une pensée pour tous ceux que j'ai rencontré en chemin, lors de conférences et écoles d'été ; Laure, Maud, Vincent, Benjamin, Andrés, Ilaria, Nelo, Clément, Lucie, Sarah, Sébastien, Cyril, Samuel et tous les autres. Merci pour tous vos conseils, et bonne chance pour la suite.
J'en profite également pour remercier la "team des agrégés" qui m'ont aidée à décompresser grâce à nos petites retrouvailles à l'Union, dans le vieux ou à Toulon.

Je termine ces remerciements par tous les membres de ma famille, qui ont toujours cru en moi. Un grand merci à mes parents pour m'avoir accompagnée à la fois dans les moments de doutes et les moments merveilleux que constituent une thèse. Une grande pensée à mon frère qui a toujours été là pour moi.
Finalement, je dédie ces derniers remerciements tendrement à mon compagnon d'aventure, de thèse et de vie. Merci infiniment pour tous tes conseils très avisés, de m'avoir épaulée pendant les moments difficiles, et de m'avoir encouragée tout au long de cette thèse.

Et maintenant, place aux maths !

Contents

Introduction (français) **11**
0.1 Tests d'indépendance non-paramétriques . . . 11
0.2 Étude des performances . . . 22
0.3 Application à la détection de synchronisations en neurosciences . . . 32

Introduction (english) **47**
0.4 Non-parametric tests of independence . . . 47
0.5 How to study the performance? . . . 57
0.6 Application to synchrony detection in neuroscience . . . 67

1 Bootstrap and permutation tests of independence for point processes **81**
1.1 Introduction . . . 82
1.2 From neuroscience interpretations to general test statistics . . . 85
1.3 Bootstrap tests of independence . . . 88
1.4 Permutation tests of independence . . . 96
1.5 Simulation study . . . 98
1.6 Conclusion . . . 102
1.7 Proofs . . . 103
1.8 Additional Results . . . 137

2 A distribution free UE method based on delayed coincidence count **141**
2.1 Introduction . . . 142
2.2 Binned and delayed coincidence counts . . . 143
2.3 Some distribution free independence tests . . . 146
2.4 Multiple tests . . . 159
2.5 Conclusion . . . 162

3 Concentration inequalities for randomly permuted sums **165**
3.1 Introduction and motivation . . . 166
3.2 Bernstein-type concentration inequalities for permuted sums . . . 172
3.3 Proofs . . . 175
3.4 Appendix: a non-asymptotic control of the second kind error . . . 184

4 Separation rates for independence tests based on wavelets **195**
4.1 Introduction . . . 196
4.2 Description of the permutation testing procedure . . . 199
4.3 Theoretical Results . . . 202
4.4 Conclusion . . . 207

4.5 Proofs 208
4.6 Appendix 215

Conclusion and Perspectives (english) **217**

Conclusion et Perspectives (français) **221**

A Mathematical Tools **225**
A.1 Some convergence results 225
A.2 The point process setting and the Skorohod topology 229
A.3 Some basics on U-statistics 231
A.4 Wavelets and Besov spaces 234

Bibliography **239**

Développement (français) **251**

Notation

Throughout the thesis

$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$	the sets of natural numbers, of integers and of real numbers.
$\lfloor\cdot\rfloor$, $\lceil\cdot\rceil$	the floor and the ceiling functions.
$\mathbb{1}_A$	the indicator function of the set A ($\mathbb{1}_A(u) = 1$ if $u \in A$ and 0 otherwise).
δ_y	the delta measure concentrated at y ($\forall$ measurable f, $\int f(t)\delta_y(dt) = f(y)$).
$\mathcal{N}(m, v)$	the standard Gaussian distribution with mean m and variance v.
$\Phi_{m,v}$, $\Phi^{-1}_{m,v}$	the cumulative distribution function and the quantile function of $\mathcal{N}(m, v)$.

$(\Omega, \mathcal{A}, \mathbb{P})$	a probability space.
$\mathbb{E}[Y]$, $\mathrm{Var}(Y)$	the expectation and the variance of a real random variable Y.
$(\mathcal{X}^2, d^2_{\mathcal{X}})$, $\mathcal{B}_{\mathcal{X}^2}$	a (product) metric space and the corresponding σ-algebra of Borel sets.
d	the product metric on $\mathcal{X}^2 \times \mathcal{X}^2$.
P, P^1, P^2	a probability measure on $\mathcal{X}^2$ and its marginals.
$\mathbb{X}_n$	a sample of n i.i.d. r.v. $X_i = (X_i^1, X_i^2)$ in $\mathcal{X}^2$ with distribution P.
P_n, P_n^1, P_n^2	the empirical probability measure associated to $\mathbb{X}_n$, the empirical marginals.
$\mathbb{X}_n^*$	the bootstrap sample associated to $\mathbb{X}_n$, with distribution $P_n^1 \otimes P_n^2$.
$\mathbb{E}^*[\cdot]$	the conditional expectation given $\mathbb{X}_n$.
$\mathfrak{S}_n$	the set of all permutations of $\{1, 2, \ldots, n\}$.
π_n, Π_n	a deterministic and a random permutation (independent on $\mathbb{X}_n$) in $\mathfrak{S}_n$.
$\mathbb{X}_n^\star$	the bootstrap sample associated to $\mathbb{X}_n$, with distribution $P_n^\star$.

Notice the difference between the star "$*$" representing the bootstrapped quantities, and the star "$\star$" representing the permuted quantities all along this manuscript.

For any functional $Z : (\mathcal{X}^2)^n \to \mathbb{R}$,

$\mathcal{L}(Z, P)$	the distribution of $Z(\mathbb{X}_n)$, where $\mathbb{X}_n$ i.i.d. sample from P.
$\mathcal{L}\left(Z, P_n^1 \otimes P_n^2 \middle\vert \mathbb{X}_n\right)$	the conditional distribution of the bootstrapped statistic $Z(\mathbb{X}_n^*)$ given $\mathbb{X}_n$.
$\mathcal{L}(Z, P_n^\star \vert \mathbb{X}_n)$	the conditional distribution of the permuted statistic $Z(\mathbb{X}_n^\star)$ given $\mathbb{X}_n$.

$\xrightarrow{a.s.}$, $\xrightarrow{\mathbb{P}}$, $\xrightarrow{\mathcal{L}}$	convergence almost sure, in probability and in distribution.
$Q_n \underset{n\to+\infty}{\Longrightarrow} Q$	weak convergence of measures Q_n to Q.
d_{BL}	the bounded Lipschitz metric.
d_2	the $\mathbb{L}_2$-Wasserstein metric, also called the Mallows metric.

In the density framework

f	a density f on $[0,1]^2$.
$f_1 \otimes f_2$	the product of the marginals of f, that is $f_1 \otimes f_2(x^1, x^2) = f_1(x^1) f_2(x^2)$.
P_f	the distribution with density f w.r.t. the Lebesgue measure.
$\mathbb{L}_p(I),\ I \subset \mathbb{R}^d$	the set of measurable functions $g : I \to \mathbb{R}$ such that $\int_I \lvert g(x)\rvert^p \, dx < +\infty$.
$\lVert \cdot \rVert_p$	the $\mathbb{L}_p$-norm defined for all g in $\mathbb{L}_p(I)$ by $\sup_{x \in I} \lvert g(x)\rvert$ if $p = \infty$ and $\lVert g\rVert_p = \left(\int_I \lvert g(x)\rvert^p \, dx\right)^{1/p}$ if $1 \leq p < \infty$.

Introduction (français)

Contents

0.1 Tests d'indépendance non-paramétriques 11
0.1.1 Deux approches non-paramétriques : le bootstrap et la permutation 11
0.1.2 Tests d'hypothèses ; le cadre mathématique 14
0.1.3 Construction de deux tests d'indépendance non-paramétriques . . . 18
0.2 Étude des performances 22
0.2.1 Étude asymptotique 22
0.2.2 Étude non-asymptotique 26
0.2.3 Inégalités de concentration : un outil performant 29
0.3 Application à la détection de synchronisations en neurosciences 32
0.3.1 Contexte biologique et motivation 32
0.3.2 Modélisation par des processus ponctuels 37
0.3.3 Une brève introduction aux tests multiples 41
0.3.4 La méthode Permutation UE et application à de vraies données . . . 42

0.1 Tests d'indépendance non-paramétriques

Cette thèse s'inscrit dans le vaste domaine des tests d'indépendance non-paramétriques. Tester l'indépendance est l'une des questions centrales de l'analyse de données et a donc été largement étudié dans la littérature en statistique. Cette thèse est motivée par des problématiques en neurosciences qui sont entièrement décrites en Section 0.3 et pour lesquelles aucun modèle n'est communément admis à ce jour. Par conséquent, nous nous focalisons sur les tests ne dépendant pas de la distribution des observations, appelés tests *non-paramétriques*.

0.1.1 Deux approches non-paramétriques : le bootstrap et la permutation

De nombreuses méthodes ont été développées pour construire des tests non-paramétriques d'indépendance. Parmi celles-ci, une première approche "naïve" consiste à définir une statistique de test dont la loi asymptotique ne dépend d'aucun paramètre inconnu sous l'hypothèse nulle. Par exemple, nous pouvons considérer le test d'indépendance historique du chi-deux de Pearson [131, 132], dont la loi asymptotique est, comme son nom l'indique, une loi du chi-deux. Ainsi, étant donné un niveau α dans $]0,1[$, nous pouvons aisément construire un test de nature purement asymptotique rejetant l'hypothèse nulle lorsque la statistique de test est strictement supérieure au quantile d'ordre $(1-\alpha)$ de la loi asymptotique, et ainsi obtenir un test ayant de bonnes propriétés asymptotiques. Cependant, dans de nombreux domaines

d'application tels que la biologie (comme notamment dans la section 0.3.1), peu d'observations sont disponibles, pour des raisons économiques ou biologiques, et de tels procédures purement asymptotiques ne sont pas appropriées.

Une autre famille importante de tests d'indépendance non-paramétriques est celle des *tests de rang*, basés sur le rang des données dans l'échantillon. Parmi ces méthodes, nous pouvons considérer la méthode de Pitman [140], basée sur le coefficient de corrélation de Pearson, celles de Hotelling et Pabst [89] ou de Kendall et al. [105], basées sur les coefficient de corrélation des rangs de Spearman, celle de Kendall [104] basée sur le tau de Kendall, ou encore celle de Wolfowitz [179] basée sur un rapport de vraisemblances adapté. Toutes ces méthodes dépendent fortement d'une relation d'ordre, et sont construites pour des variables ou des vecteurs aléatoires réels. Cependant, en raison de la motivation biologique détaillée en section 0.3.1, nous souhaitons construire des tests d'indépendance pouvant s'appliquer à des variables aléatoires plus générales, et en particulier à des processus ponctuels, pour lesquels il n'existe aucune relation d'ordre naturelle. Ainsi, nous nous sommes tournés vers des méthodes applicables dans des cas plus généraux, basées sur des approches par bootstrap ou par permutation.

L'approche par permutation : un concept pionnier

Également appelés tests par randomisation, les tests par permutation ont été introduits par Fisher [54] en 1935 afin d'éprouver l'affirmation d'une dame prétendant pouvoir déterminer lequel du lait ou du thé a été versé en premier dans la tasse. Fisher construit une expérience permettant de décider si l'hypothèse nulle, à savoir ($\mathcal{H}_0$) "la dame n'a pas de discernement sensoriel", est significativement fausse ou non. Basé sur une permutation aléatoire de différentes tasses dans lesquelles le lait a été versé soit avant, soit après le thé, il calcule les probabilités sous l'hypothèse nulle de toutes les différentes erreurs que la dame pourrait commettre en répondant complètement au hasard. Cela lui permet donc de construire un test valide, contrôlant la probabilité de rejeter à tort ($\mathcal{H}_0$), c'est-à-dire de croire l'affirmation de la dame alors qu'en réalité, elle ment. L'argument principal validant cette approche est que, sous l'hypothèse nulle, les tasses sont "échangeables" dans le sens que leur goût est le même pour la dame. Cette idée est formalisée et généralisée ci-dessous.

L'hypothèse fondamentale suivante, énoncée dans l'article de Hoeffding [84], est à la base des tests par permutation/randomisation :

($\mathcal{A}$) L'ensemble des mesures de probabilité vérifiant l'hypothèse nulle est invariant sous l'action d'un groupe de transformations G.

Cela signifie que, si un échantillon $\mathbb{Y}_n = (Y_1, \ldots, Y_n)$ est tiré aléatoirement selon une loi vérifiant l'hypothèse nulle (par exemple, les m premières variables $Y_1, \ldots, Y_m$ ont la même loi, ainsi que les $(n-m)$ dernières $Y_{m+1}, \ldots, Y_n$), alors, appliquer n'importe quelle transformation du groupe G (par exemple, les $m!(n-m)!$ permutations des m premières et des $(n-m)$ dernières variables) ne change pas la loi. En d'autre termes, pour toute transformation g dans G, $g\mathbb{Y}_n$ et $\mathbb{Y}_n$ ont la même loi. En particulier, cela fournit une manière simple de mimer la loi de la statistique de test sous l'hypothèse nulle, et ainsi de construire des valeurs critiques valides. Remarquons que, généralement, si l'échantillon $\mathbb{Y}_n$ ne vérifie pas l'hypothèse nulle, l'affirmation précédente n'est plus vraie, et plus aucun contrôle sur la loi de l'échantillon transformé n'est garanti. Cependant, dans le cas particulier des tests d'indépendance, cette approche (lorsqu'elle est appliquée correctement) permet de reconstruire la loi sous l'hypothèse

nulle, et ce, même si l'observation $\mathbb{Y}_n$ ne vérifie pas l'hypothèse nulle. En particulier, c'est le cas pour les tests par permutation construits dans cette thèse.

Depuis les travaux précurseurs de Fisher, les tests par permutation ont été largement développés dans de nombreux contextes, notamment par Pitman dans sa série d'articles [139, 140, 141]. Le lecteur intéressé peut se référer aux les livres de Pesarin et Salmaso [135], d'Efron et Tibshirani [51] ou de Good [63] pour plus de détails sur de telles méthodes, et [4, 36, 88, 100, 111, 112] pour des travaux plus récents. Dans le cadre particulier des tests d'indépendance, les articles de Scheffe [161], ou de Hoeffding [84] sont dans la lignée des tests de rang cités ci-dessus. Dans cette thèse, nous nous intéressons particulièrement aux tests par permutation de type Kolmogorov-Smirnov de Romano [155] ou de van der Vaart et Wellner [173]. Plus récemment, Gretton et ses co-auteurs [65, 66] ont également appliqué l'approche par permutation afin d'approcher la loi sous l'hypothèse nulle de leur statistique de test basée sur des noyaux reproduisants.
Un lecteur attentif remarquera un vide dans la littérature entre les années 1950 et les années 1990. Une des raisons principales pour lesquelles l'approche par permutation a été laissée de coté pendant cette période est due à son énorme coût en termes de temps de calculs. Cependant, de nos jours, elle a regagnée en popularité d'une part grâce aux progrès énormes en informatique, et d'autre part grâce aux justifications théoriques de l'utilisation d'approximations par Monte Carlo (voir par exemple le lemme de Romano et Wolf [156, Lemme 1]).

De la randomisation au bootstrap

Généralement, la justification des approches par permutation repose sur les probabilités conditionnelles sachant l'observation $\mathbb{Y}_n$, évitant ainsi la nécessité de faire des hypothèses souvent restrictives sur la distribution de cette dernière. Cette idée de travailler conditionnellement à $\mathbb{Y}_n$, d'abord développée pour les tests, a été généralisée à d'autres domaines de la statistique tels que l'estimation ou la construction d'intervalles de confiance. Dans la lignée de ces travaux, Quenouille [144] a développé la méthode du *Jackknife* pour estimer le biais de certains estimateurs sans avoir à faire d'hypothèse sur la loi sous-jacente des données. Ensuite, l'approche par bootstrap, introduite par Efron [50] à la fin des années 1970, généralise et améliore le Jackknife.

L'idée principale du bootstrap d'Efron est la suivante. Considérons un échantillon $\mathbb{Y}_n = (Y_1, \dots, Y_n)$ de loi inconnue P, et supposons que nous voulions estimer la loi empirique de la variable aléatoire $\bar{R}_n = R_n(\mathbb{Y}_n, P)$ dépendant de la donnée $\mathbb{Y}_n$ ainsi que de sa loi sous-jacente P. À titre d'exemple, nous pouvons considérer, comme dans [50], le biais de la moyenne empirique défini comme $R_n(\mathbb{Y}_n, P) = n^{-1}\sum_{i=1}^n Y_i - \int y dP(y)$. L'idée du bootstrap est alors de remplacer la vraie loi P par sa version empirique $P_n = n^{-1}\sum_{i=1}^n \delta_{Y_i}$, où δ_y désigne la masse de Dirac au point y, et ce à tous les niveaux. Plus précisément, considérons d'abord l'échantillon bootstrappé $\mathbb{Y}_n^*$ constitué de n copies indépendantes et identiquement distribuées (i.i.d.) de loi P_n, qui sont obtenues concrètement en tirant avec remise dans l'échantillon initial $\mathbb{Y}_n$. Ensuite, la loi de $\bar{R}_n$ est estimée par la loi conditionnelle de $R_n^* = R_n(\mathbb{Y}_n^*, P_n)$ sachant $\mathbb{Y}_n$. Remarquons que cette loi conditionnelle est totalement connue une fois l'échantillon $\mathbb{Y}_n$ donné, et prend au plus n^n valeurs obtenues à partir du rééchantillonnage avec remise.

Selon l'heuristique d'Efron, la loi conditionnelle de R_n^* sachant $\mathbb{Y}_n$ est proche de celle de $\bar{R}_n$, la plus grande difficulté résidant dans sa justification. Cette méthode, comme la permutation, est largement appliquée à de nombreuses statistiques et étudiée dans la littérature (voir le livre d'Efron et Tibshirani [51] pour un état des lieux complet). De nombreux résultats de type bootstrap, prouvant que la loi conditionnelle de R_n^* sachant $\mathbb{Y}_n$ converge presque sûrement

ou en probabilité vers la loi asymptotique de $\bar{R}_n$ ont été obtenus, non seulement pour le bootstrap de la moyenne, comme décrit ci-dessus [20], mais aussi, par exemple, par Arcones et Giné [5] pour les U-statistiques ou par Bickel et Freedman [20] pour des fonctionnelles de von Mises. De plus, il faut garder à l'esprit que toutes les approches par bootstrap qui ont été validées théoriquement dans la littérature concernent des statistiques centrées (c'est-à-dire que $\mathbb{E}\left[\bar{R}_n\right] = 0$). L'importance de ce centrage est illustré dans le chapitre 2.

L'approche d'Efron a inspiré de nombreuses généralisations, telles que le bootstrap à poids. Avec les mêmes notations que ci-dessus, considérons pour tout i appartenant à $\{1, \dots, n\}$, la variable aléatoire $M_{n,i} = \sum_{j=1}^{n} \mathbb{1}_{Y_j^* = Y_i}$ comptant le nombre de fois que la $i$ème variable Y_i apparaît dans l'échantillon bootstrappé $\mathbb{Y}_n^*$. La mesure empirique de ce dernier peut alors être exprimée en fonction de l'échantillon initial, en écrivant $P_n^* = n^{-1} \sum_{i=1}^{n} \delta_{Y_i^*} = n^{-1} \sum_{i=1}^{n} M_{n,i} \delta_{Y_i}$. L'idée du bootstrap à poids, pour lequel Præstgaard et Wellner [143] ont été les premiers à obtenir des résultats généraux, est de remplacer les $M_{n,i}$ par des poids échangeables $W_{n,i}$. Par exemple, nous pouvons considérer des poids positifs déterministes $w_{n,i}$ tels que $\sum_{i=1}^{n} w_{n,i} = n$, une permutation aléatoire Π_n de loi uniforme sur l'ensemble $\mathfrak{S}_n$ de toutes les permutations de $\{1, \dots, n\}$, et définir les nouveaux poids par $W_{n,i} = w_{n,\Pi_n(i)}$. La mesure empirique associée s'écrit donc $P_n^W = n^{-1} \sum_{i=1}^{n} w_{n,\Pi_n(i)} \delta_{Y_i}$. En particulier, en prenant $w_{n,1}$ égal à 0, et $w_{n,i}$ égal à $n/(n-1)$ pour tout i dans $\{2, \dots, n\}$, nous retrouvons l'idée du Jackknife. D'autres choix de poids mènent à d'autres types de méthodes par bootstrap telles que le *m out of n bootstrap* de Bretagnolle [30], ou le bootstrap sauvage basé sur des poids indépendants. D'autres aspects du bootstrap sont décrits dans le cours de Giné [62].

De nombreux tests d'indépendance basés sur des approches par bootstrap ont été étudiés, tels que par exemple, ceux de type Kolmogorov-Smirnov de Romano [154] ou de van der Vaart et Wellner [173], rappelés dans la section 0.1.3.

La permutation apparaît parfois dans la littérature comme un cas particulier du bootstrap en tant que méthode de rééchantillonnage. Cependant, elle consiste en un tirage *sans* remise, alors que le bootstrap en est un *avec* remise. Dans la suite, nous distinguerons toujours les deux méthodes puisqu'elles sont très différentes par nature, et que leurs études sont basées sur des approches complètement différentes, comme nous pouvons le constater dans le chapitre 1. Par ailleurs, ces deux approches (la permutation et le bootstrap) fonctionnent conditionnellement aux données et diffèrent donc des procédures de test usuelles dans le sens où les valeurs critiques sont aléatoires, puisqu'elles dépendent des données. De plus, étant donné que, pour de grandes valeurs de n (telles que $n \geq 15$), le calcul de tous les échantillons permutés ou bootstrappés est très coûteux en temps de calcul (respectivement au moins $n!$ ou n^n), des méthodes de Monte Carlo sont en général appliquées en pratique.

Les tests d'indépendance non-paramétriques présentés et étudiés dans cette thèse sont essentiellement inspirés de ceux par permutation et par bootstrap de van der Vaart et Wellner [173] ainsi que de Romano [155], lui-même inspiré par les tests par permutation de Hoeffding [84]. Un des premiers objectifs de ces travaux est d'adapter et de généraliser leur statistique de tests afin de traiter les problématiques en neurosciences présentées dans la section 0.3.1.

0.1.2 Tests d'hypothèses ; le cadre mathématique

Tout au long de cette thèse, les notations suivantes sont utilisées. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(\mathcal{X}, d_{\mathcal{X}})$ un espace métrique. Considérons $X = (X^1, X^2)$ une variable aléatoire sur $\mathcal{X}^2$, c'est-à-dire une fonction mesurable de $(\Omega, \mathcal{A})$ dans $(\mathcal{X}^2, \mathcal{B}_{\mathcal{X}^2})$, où $\mathcal{B}_{\mathcal{X}^2}$ désigne la tribu borélienne sur $(\mathcal{X}^2, d_{\mathcal{X}^2})$ et $d_{\mathcal{X}^2}$ est une mesure produit issue de $d_{\mathcal{X}}$. La variable aléatoire

X est dite de loi P et de marginales P^1 et P^2 si pour tout ensemble borélien B de $\mathcal{X}^2$, $\mathbb{P}(X \in B) = P(B)$, et pour tous ensembles boréliens B^1 et B^2 de $\mathcal{X}$, $\mathbb{P}\big(X^1 \in B^1\big) = P^1(B^1)$ et $\mathbb{P}\big(X^2 \in B^2\big) = P^2(B^2)$. Le produit des marginales de P, noté $P^1 \otimes P^2$, est alors défini comme $P^1 \otimes P^2(B^1 \times B^2) = P^1(B^1)P^2(B^2)$, et les variables X^1 et X^2 sont indépendantes si et seulement si $P = P^1 \otimes P^2$.

Deux ensembles d'observations. Le premier cadre, principalement considéré dans les chapitres 1 et 2, est l'ensemble des processus ponctuels. Il est adapté à l'application en neurosciences décrite dans la section 0.3. Dans ce cas, $\mathcal{X}$ représente l'ensemble de toutes les sous-ensembles dénombrables de $[0,1]$, muni de la métrique issue de la topologie de Skorokhod décrite en annexe A.2.2.
Le second cadre, principalement considéré dans le chapitre 4, est celui plus classique des variables à densité, où $\mathcal{X}$ désigne un intervalle réel compact, ici égal à $[0,1]$ sans perte de généralités, et les variables aléatoires sur $\mathcal{X}^2$ sont supposées être à densité dans $\mathbb{L}_2(\mathcal{X}^2)$ par rapport à la mesure de Lebesgue.

Les tests d'indépendance ; la problématique. Dans la suite, $\mathbb{X}_n = (X_1, \dots, X_n)$ désigne un échantillon de n copies i.i.d. $X_i = (X_i^1, X_i^2)$ (pour $1 \leq i \leq n$) d'une variable aléatoire $X = (X^1, X^2)$ sur $\mathcal{X}^2$, de loi P et de marginales P^1 et P^2. Nous l'appellerons observation, échantillon (observé) ou encore données initiales tout au long de cette thèse. Étant donnée l'observation d'un échantillon $\mathbb{X}_n$, nous souhaitons tester l'hypothèse nulle $(\mathcal{H}_0)$ "les coordonnées X^1 et X^2 de X sont indépendantes" contre l'alternative $(\mathcal{H}_1)$ "elles ne le sont pas", ce qui équivaut à tester

$$(\mathcal{H}_0)\ P = P^1 \otimes P^2 \quad \text{contre} \quad (\mathcal{H}_1)\ P \neq P^1 \otimes P^1.$$

Dans la suite, un test statistique de $(\mathcal{H}_0)$ contre $(\mathcal{H}_1)$ est une fonction Δ de l'échantillon $\mathbb{X}_n$, éventuellement dépendant du nombre d'observations n, à valeurs dans $\{0,1\}$, $\Delta(\mathbb{X}_n) = 1$ signifiant que l'hypothèse nulle $(\mathcal{H}_0)$ est rejetée au profit de l'alternative $(\mathcal{H}_1)$, $\Delta(\mathbb{X}_n) = 0$ signifiant que $(\mathcal{H}_0)$ ne peut être rejetée au vue des données, et de ce fait, est acceptée. Un test statistique peut avoir deux types d'erreurs. L'*erreur de première espèce*, aussi appelée faux positif, revient à rejeter l'hypothèse nulle à tort, et l'*erreur de seconde espèce*, aussi appelée faux négatif, revient à l'accepter à tort. Notons $\mathcal{P}_0$ (respectivement $\mathcal{P}_1$) l'ensemble des mesures de probabilité sur $(\mathcal{X}^2, \mathcal{B}_{\mathcal{X}^2})$ vérifiant $(\mathcal{H}_0)$ (respectivement $(\mathcal{H}_1)$), c'est-à-dire étant égales au (respectivement différentes du) produit de leurs marginales. Les *risques de première et seconde espèces* sont alors définis comme

$$\sup_{P \in \mathcal{P}_0} \mathbb{P}_P(\Delta(\mathbb{X}_n) = 1) \quad \text{et} \quad \sup_{P \in \mathcal{P}_1} \mathbb{P}_P(\Delta(\mathbb{X}_n) = 0)\,,$$

où l'indice P dans $\mathbb{P}_P(\Delta(\mathbb{X}_n) = 0)$ signifie que $\mathbb{X}_n$ un échantillon de variables aléatoires i.i.d. de loi P.
Pour tout niveau prescrit α dans $]0,1[$, nous dirons qu'un test Δ est *de niveau exact* α si son risque de première espèce est contrôlé par α, et ce, quelle que soit la taille de l'échantillon n. De plus, nous dirons qu'il est *asymptotiquement de taille* α si son risque de première espèce converge vers α lorsque le nombre d'observations n tend vers $+\infty$.
La *puissance* du test est définie comme la fonction $P \in \mathcal{P}_1 \mapsto 1 - \mathbb{P}_P(\Delta(\mathbb{X}_n) = 0)$. Finalement, nous dirons qu'un test statistique est *consistent contre une alternative* P dans $\mathcal{P}_1$ si sa puissance, contre cette alternative, converge vers 1 lorsque n tend vers $+\infty$.

Construction d'un test statistique. Étant donnée l'observation d'un échantillon $\mathbb{X}_n$ de variables aléatoire i.i.d. de loi inconnue P, il existe plusieurs manières de construire un test statistique au niveau prescrit α dans $]0,1[$. Toutes les procédures de test commencent par la définition d'une *statistique de test* T à valeurs réelles, c'est-à-dire une fonction mesurable de $\mathbb{X}_n$ bien adaptée aux hypothèses à tester. La connaissance de sa loi sous l'hypothèse nulle, à savoir lorsque P vérifie $(\mathcal{H}_0)$, est l'objectif central lors de la construction d'un test statistique.

Ensuite, l'approche classique consiste à définir une *région critique* R_α (dépendant de α) incluse dans $\mathbb{R}$. Le test est alors défini comme $\Delta_\alpha(\mathbb{X}_n) = \mathbb{1}_{T(\mathbb{X}_n)\in R_\alpha}$. En général, la région critique R_α est établie de manière à ce que le test soit de niveau exact α (si possible), ou alors asymptotiquement de taille α. Nous pouvons alors distinguer trois types de tests : les tests *unilatéraux à droite*, *unilatéraux à gauche* ou *bilatéraux*, pour lesquels les régions critiques sont respectivement de la forme $]c_\alpha^+, +\infty[$, $]-\infty, c_\alpha^-[$ et $]-\infty, c_{\alpha/2}^-[\cup]c_{\alpha/2}^+, +\infty[$. Précisons que si la loi sous l'hypothèse nulle, ou sa limite lorsque n tend vers $+\infty$, est connue, alors les valeurs critiques c_α^+ et c_α^- sont choisies comme étant les quantiles d'ordre $(1-\alpha)$ et α de cette loi.

Une autre approche est basée sur la notion de *p-valeur* (également appelée *niveau de significativité* par Efron et Tibshirani [51]). La p-valeur correspondant à une procédure de test $\mathbf{\Delta} = \{\Delta_\alpha\}_{\alpha\in(0,1)}$, introduite comme ci-dessus, est définie comme $\hat{\alpha}(\mathbb{X}_n) = \sup\{\alpha'; \Delta_{\alpha'}(\mathbb{X}_n) = 0\}$. Elle correspond au plus grand niveau de significativité α' sous lequel le test $\Delta_{\alpha'}$ accepte l'hypothèse nulle au vue de l'observation $\mathbb{X}_n$. Ainsi, un test de niveau prescrit α peut être naturellement défini comme $\Delta_\alpha(\mathbb{X}_n) = \mathbb{1}_{\hat{\alpha}(\mathbb{X}_n)\leq\alpha}$. Un tel test est alors de n'importe quel niveau exact si quelque soit u dans $]0,1[$, $\mathbb{P}(\hat{\alpha}(\mathbb{X}_n) \leq u) \leq u$ sous l'hypothèse nulle. Intuitivement, la p-valeur représente la crédibilité de l'hypothèse nulle au vue de l'observation. Plus elle est petite, moins l'hypothèse nulle $(\mathcal{H}_0)$ est fiable. En particulier, comme le mentionnent Efron et Tibshirani dans [51], il est communément admis qu'une p-valeur $\hat{\alpha}(\mathbb{X}_n)$ inférieure à 0.1, 0.05, 0.025 et 0.01 correspondent respectivement à des arguments "limites", "relativement forts", "forts" et "très forts" contre $(\mathcal{H}_0)$. Lorsque la loi sous l'hypothèse nulle de la statistique de test $T(\mathbb{X}_n)$, ou sa limite lorsque le nombre d'observations n tend vers $+\infty$, est connue, la p-valeur peut alors s'écrire en fonction de la version continue à gauche de sa fonction de répartition, notée F_0^-. Par exemple, la p-valeur correspondant au test unilatéral à droite vérifie $\hat{\alpha}(\mathbb{X}_n) = 1 - F_0^-(T(\mathbb{X}_n))$. En particulier, elle est égale à la probabilité sous l'hypothèse nulle d'obtenir une statistique de test supérieure ou égale à la valeur observée $T(\mathbb{X}_n)$, ou plus précisément, $\mathbb{P}\big(T(\mathbb{X}_n^0) \geq T(\mathbb{X}_n)|\mathbb{X}_n\big)$ où $\mathbb{X}_n^0$ est un échantillon indépendant de $\mathbb{X}_n$ et de même loi (sous $(\mathcal{H}_0)$).
Dans cette thèse, les deux constructions sont considérées (voir la section 0.1.3) : les valeurs critiques sont définies dans le chapitre 1 et sont traduites en termes de p-valeurs dans le chapitre 2.

Optimalité au sens du minimax. Généralement, il existe plusieurs procédures de test pour un même problème à tester. Il est donc naturel de se demander comment comparer deux tests, et en particulier, comment décrire les performances d'un test. Dans cette thèse, les performances non-asymptotiques d'un test sont exprimées en termes de vitesses de séparation uniforme, décrites ci-dessous.
Étant donnés deux niveaux de risques α et β dans $]0,1[$, nous souhaitons construire un test de $(\mathcal{H}_0)$ "$P \in \mathcal{P}_0$" étant de niveau exact α, et ayant un risque de seconde espèce contrôlé par β. Habituellement, la région critique est ajustée de manière à ce que le risque de première espèce soit automatiquement contrôlé par α. Puisqu'il ne reste aucune variabilité permettant d'imposer un contrôle du risque de seconde espèce, l'idée est de restreindre l'ensemble des

alternatives sur lequel le supremum est pris dans la définition de ce risque.
Intuitivement, plus l'alternative est loin de l'hypothèse nulle, plus le test est capable de rejeter cette dernière. L'idée est donc d'introduire l'ensemble des alternatives qui sont à une distance supérieure à $\rho > 0$ de l'hypothèse nulle. Cependant, ce sous-ensemble d'alternatives demeurant souvent trop grand, nous le restreignons encore à une famille $\{\mathcal{Q}_\nu\}_{\nu\in\mathcal{M}}$ de sous-espaces de $\mathcal{P}$ ayant des propriétés de régularité. Par exemple, ainsi que dans le cadre de variables à densité considéré dans le chapitre 4, chaque $\mathcal{Q}_\nu$ peut être défini comme la partie de l'ensemble des mesures de probabilité ayant une densité admettant une certaine régularité de paramètre ν.

Definition 0.1.1. La *vitesse de séparation uniforme* $\rho(\Delta_\alpha, \mathcal{Q}_\nu, \beta)$ d'un test Δ_α de niveau exact α, sur une famille $\mathcal{Q}_\nu$ de lois de régularité ν par rapport à une distance $\bar{d}$ sur l'ensemble des mesures de probabilité de $\mathcal{X}^2$ est définie par

$$\rho(\Delta_\alpha, \mathcal{Q}_\nu, \beta) = \inf\left\{\rho > 0\ ;\ \sup_{P\in\mathcal{Q}_\nu;\ \bar{d}(P,\mathcal{P}_0)>\rho} \mathbb{P}_P(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta\right\}.$$

Intuitivement, cela correspond à la plus petite distance à l'hypothèse nulle à partir de laquelle le test détecte les alternatives de régularité ν avec un risque de seconde espèce contrôlé par β. Ainsi, lors de la comparaison de plusieurs tests de $(\mathcal{H}_0)$ "$P \in \mathcal{P}_0$" étant tous de niveau exact α, celui ayant la plus petite vitesse de séparation uniforme est le plus performant puisqu'il est capable de détecter plus d'alternatives que les autres. L'idée naturelle pour décrire la performance d'un test (de niveau exact α) est donc de le comparer au "meilleur" en termes de vitesse de séparation uniforme. Pour cela, nous considérons la *vitesse de séparation minimax* de test de $(\mathcal{H}_0)$ "$P \in \mathcal{P}_0$", introduite par Baraud [13], et définie par

$$\rho(\mathcal{Q}_\nu, \alpha, \beta) = \inf\left\{\rho(\Delta_\alpha, \mathcal{Q}_\nu, \beta)\right\},$$

où l'infimum est pris sur tous les tests Δ_α de $(\mathcal{H}_0)$ qui sont de niveau exact α. Un test de niveau exact α atteignant la vitesse de séparation minimax à constante multiplicative près est dit *optimal au sens du minimax*.

La notion d'optimalité au sens du minimax a d'abord été introduite par Ingster dans une série d'articles fondamentaux [91, 95] mais d'un point de vue asymptotique.
La *vitesse de test minimax* (asymptotique) $\rho_n(\mathcal{Q}_\nu)$ sur un espace de régularité $\mathcal{Q}_\nu$ vérifie une propriété de borne supérieure :

Pour tous $\alpha, \beta > 0$, il existe $C > 0$ et un test Δ^* tels que

$$\begin{cases} \limsup_{n\to+\infty} \sup_{P\in\mathcal{P}_0} \mathbb{P}_P(\Delta^*(\mathbb{X}_n) = 1) \leq \alpha, \\ \limsup_{n\to+\infty} \sup_{P\in\mathcal{Q}_\nu,\ \bar{d}(P,\mathcal{P}_0)>C\rho_n(\mathcal{Q}_\nu)} \mathbb{P}_P(\Delta^*(\mathbb{X}_n) = 0) \leq \beta. \end{cases} \tag{0.1.1}$$

et une propriété de borne inférieure :

Pour tout ρ'_n tel que $\rho'_n/\rho_n(\mathcal{Q}_\nu) \underset{n\to+\infty}{\longrightarrow} 0$,

$$\inf_\Delta \left\{\sup_{P\in\mathcal{P}_0} \mathbb{P}_P(\Delta(\mathbb{X}_n) = 1) + \sup_{P\in\mathcal{Q}_\nu,\ \bar{d}(P,\mathcal{P}_0)>\rho'_n} \mathbb{P}_P(\Delta(\mathbb{X}_n) = 0)\right\} \underset{n\to+\infty}{\longrightarrow} 1. \tag{0.1.2}$$

Intuitivement, la borne supérieure garantit que la vitesse de test minimax est bien atteinte à constante près, et la borne inférieure s'assure que c'est la meilleure possible. Dans la littérature, la *vitesse de séparation minimax* d'un test est souvent équivalente à sa *vitesse de test*

minimax (asymptotique). Cependant, étant donné que nous souhaitons obtenir des résultats non-asymptotiques dans la suite, nous nous concentrons sur la première notion.

Si, en plus d'être optimal au sens du minimax sur $\mathcal{Q}_\nu$, un test ne dépend pas du paramètre de régularité ν, il est alors dit être *adaptatif au sens du minimax*. Un coût en $\ln(n)$ ou $\ln(\ln(n))$ est parfois inévitable pour l'adaptativité, selon la classe d'alternatives $\mathcal{Q}_\nu$ considérée (voir Spokoiny [167]).

0.1.3 Construction de deux tests d'indépendance non-paramétriques

Par souci de clarté, seul les tests unilatéraux à droite sont présentées dans cette introduction, mais les tests unilatéraux à gauche et ceux bilatéraux sont aussi définis et étudiés dans le chapitre 1.
Supposons que nous observions un échantillon $\mathbb{X}_n$ de n variables aléatoires i.i.d. de loi P sur $\mathcal{X}$, et fixons un niveau α dans $]0,1[$.

La statistique de test

Avant de décrire comment le bootstrap et la permutation permettent de construire les valeurs critiques, commençons par introduire la statistique de test. Elle est donnée par $\sqrt{n}U_{n,h}(\mathbb{X}_n)$, où $U_{n,h}(\mathbb{X}_n)$ est une U-statistique basée sur un noyau $h : \mathcal{X}^2 \times \mathcal{X}^2 \to \mathbb{R}$ mesurable symétrique, définie par

$$U_{n,h}(\mathbb{X}_n) = \frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} h(X_i, X_j). \tag{0.1.3}$$

Supposons que le noyau vérifie l'hypothèse de recentrage suivante :

$$\int_{\mathcal{X}^2 \times \mathcal{X}^2} h(x,y) dP(x) dP(y) = 0,$$

garantissant que la statistique de test est centrée sous l'hypothèse nulle. Il est important de noter que cette hypothèse est essentielle pour que l'approche par bootstrap soit applicable, comme nous pouvons le voir dans la section 2.3 du chapitre 2. En général, les U-statistiques fournissent des outils bien adaptés aux tests non-paramétriques, en tant qu'estimateurs eux-mêmes non-paramétriques dont le comportement asymptotique a été largement étudié au cours des dernières décennies (voir par exemple [164, Chapitre 5]). Une brève introduction à de telles statistiques est présentée en annexe A.3.

Un cas particulier important considéré dans cette thèse, appelé le *cas linéaire*, consiste à prendre un noyau h de la forme h_φ défini pour tout $x = (x^1, x^2)$ et $y = (y^1, y^2)$ dans $\mathcal{X}^2$ comme

$$h_\varphi(x,y) = \frac{1}{2}\left[\varphi(x^1,x^2) + \varphi(y^1,y^2) - \varphi(x^1,y^2) - \varphi(y^1,x^2)\right], \tag{0.1.4}$$

où $\varphi : \mathcal{X}^2 \to \mathbb{R}$ est une fonction mesurable. En particulier, l'hypothèse de recentrage est automatiquement vérifiée par de tels noyaux. Nous pouvons remarquer que dans ce cas, la U-statistique est un estimateur sans biais de

$$\int_{\mathcal{X}^2} \varphi(x^1,x^2)\left[dP(x^1,x^2) - dP^1(x^1)dP^2(x^2)\right], \tag{0.1.5}$$

et ce sans aucune hypothèse sur la loi sous-jacente P de $\mathbb{X}_n$. Dans la littérature, la quantité décrite dans (0.1.5) est à la base de nombreux tests d'indépendance. En effet, pour des

fonctions φ bien choisies, ou éventuellement en prenant un supremum sur une famille de telles fonctions, cette quantité fournit une pseudo-distance entre la loi jointe P et le produit de ses marginales $P^1 \otimes P^2$. Le fait qu'elle soit égale à zéro caractérise donc l'indépendance.
Dans le cas d'un supremum, nous retrouvons les statistiques de test de type Kolmogorov-Smirnov de Romano [154, 155] ou de van der Vaart et Wellner [173], qui s'écrivent dans nos notations comme

$$H_n(\mathbb{X}_n) = \frac{(n-1)}{n} \times \sup_{(v^1,v^2)\in\mathcal{V}^1\times\mathcal{V}^2} \left|\sqrt{n}U_{n,h_{\varphi_{(v^1,v^2)}}}(\mathbb{X}_n)\right|, \tag{0.1.6}$$

où respectivement

- $\mathcal{V}^1$ et $\mathcal{V}^2$ sont des classes de Vapnik-Chervonenkis au plus dénombrables de sous-ensembles de $\mathcal{X}$, et $\varphi_{(v^1,v^2)}(x^1,x^2) = \mathbb{1}_{v^1}(x^1)\mathbb{1}_{v^2}(x^2)$,
- $\mathcal{V}^1$ et $\mathcal{V}^2$ sont des classes bien choisies de fonctions mesurables sur $\mathcal{X}$ à valeurs réelles, et $\varphi_{(v^1,v^2)}(x^1,x^2) = v^1(x^1)v^2(x^2)$.

Nous pouvons remarquer que seules des fonctions φ de type produit sont considérées dans ces travaux précédents, et à notre connaissance, dans la littérature. Cependant, en raisons de notre motivation neurobiologique, nous souhaitons étudier des formes plus générales de φ, telles que le nombre de coïncidences avec délai φ_δ^{coinc} introduit dans la section 0.3.1. Cette généralisation a notamment nécessité des développements plus approfondis, et constitue une des contributions théoriques majeures de cette thèse.

Un test d'indépendance par bootstrap

Inspirés des tests d'indépendance de Romano [154] et de van der Vaart's [173], nous avons d'abord considéré une approche par bootstrap. Au lieu de considérer l'approche "naïve" d'Efron, qui consiste à tirer avec remise dans l'échantillon initial de couples $(X_1,\dots,X_n)$, nous suivons l'idée de Romano [154] et de van der Vaart et Wellner [173] qui est de rééchantillonner selon le produit des marginales empiriques, forçant ainsi l'indépendance entre les coordonnées de l'échantillon bootstrappé. Plus précisément, considérons les marginales empiriques, notées pour $j = 1, 2$

$$P_n^j = \frac{1}{n}\sum_{i=1}^n \delta_{X_i^j}.$$

L'échantillon bootstrappé, noté $\mathbb{X}_n^* = (X_1^*,\dots,X_n^*)$, est un n-échantillon i.i.d. de loi $P_n^1 \otimes P_n^2$. Concrètement, cela consiste à tirer avec remise dans l'ensemble des premières coordonnées $\{X_1^1,\dots X_n^1\}$ et de tirer indépendamment avec remise dans l'ensemble des secondes coordonnées $\{X_1^2,\dots,X_n^2\}$. Par conséquent, conditionnellement à $\mathbb{X}_n$, les coordonnées de chaque X_i^* sont indépendantes, et l'échantillon bootstrappé $\mathbb{X}_n^*$ vérifie donc l'hypothèse nulle d'indépendance $(\mathcal{H}_0)$. Ainsi, cette approche constitue une manière simple de reconstruire la loi sous l'hypothèse nulle.

Selon le paradigme du bootstrap, la loi conditionnelle de la statistique de test bootstrappée $\sqrt{n}U_{n,h}(\mathbb{X}_n^*)$ sachant $\mathbb{X}_n$ devrait être proche de la loi de la statistique de test sous $(\mathcal{H}_0)$. Nous définissons donc comme valeur critique le quantile $q^*_{1-\alpha,n}(\mathbb{X}_n)$ d'ordre $(1-\alpha)$ de la loi conditionnelle de $\sqrt{n}U_{n,h}(\mathbb{X}_n^*)$ sachant $\mathbb{X}_n$, et le *test unilatéral à droite par bootstrap* rejette l'indépendance lorsque

$$\sqrt{n}U_{n,h}(\mathbb{X}_n) > q^*_{1-\alpha,n}(\mathbb{X}_n).$$

Remarquons que la valeur critique est aléatoire, puisqu'elle dépend de $\mathbb{X}_n$, et qu'elle peut être calculée de manière exacte en considérant les n^{2n} échantillons bootstrappés possibles. En effet, soit

$$\sqrt{n}U_{n,h}^{*(1)}(\mathbb{X}_n) \;\leq\; \ldots \;\leq\; \sqrt{n}U_{n,h}^{*(n^{2n})}(\mathbb{X}_n)$$

les valeurs ordonnées que peut prendre la statistique de test bootstrappée étant donné $\mathbb{X}_n$. Alors, la valeur critique vérifie $q_{1-\alpha,n}^{*}(\mathbb{X}_n) = \sqrt{n}U_{n,h}^{*(\lceil n^{2n}(1-\alpha)\rceil)}(\mathbb{X}_n)$.

Comme mentionné précédemment, le calcul du quantile conditionnel exact $q_{1-\alpha,n}^{*}(\mathbb{X}_n)$ est très coûteux, et ce, même pour des valeurs de n modérément grandes. Ainsi, ils sont approchés en pratique par la méthode de Monte Carlo suivante. Simulons B_n échantillons bootstrappés i.i.d. (conditionnellement à $\mathbb{X}_n$), notés $\mathbb{X}_n^{*1}, \ldots, \mathbb{X}_n^{*B_n}$. Pour tout $1 \leq b \leq B_n$, notons $U^{*b} = U_{n,h}(\mathbb{X}_n^{*b})$ et considérons la statistique d'ordre correspondante $U^{*(1)} \leq \cdots \leq U^{*(B_n)}$. Alors, le quantile approché par Monte Carlo est définie comme $q_{1-\alpha,n}^{*MC}(\mathbb{X}_n) = \sqrt{n}U^{*(\lceil B_n(1-\alpha)\rceil)}$, et le *test unilatéral à droite par bootstrap avec approximation par Monte Carlo* rejette l'indépendance lorsque

$$\sqrt{n}U_{n,h}(\mathbb{X}_n) > q_{1-\alpha,n}^{*MC}(\mathbb{X}_n).$$

La p-valeur correspondante de ce test est définie comme

$$\frac{1}{B_n}\sum_{b=1}^{B_n} \mathbb{1}_{U^{*b}\geq U_{n,h}(\mathbb{X}_n)}.$$

Remarquons que le *test unilatéral à droite par bootstrap avec approximation par Monte Carlo* est exactement le test qui rejette l'indépendance lorsque cette p-valeur est inférieure ou égale à α. En effet,

$$\begin{aligned}
\sqrt{n}U_{n,h}(\mathbb{X}_n) > q_{1-\alpha,n}^{*MC}(\mathbb{X}_n) \quad &\Leftrightarrow \quad U_{n,h}(\mathbb{X}_n) > U^{*(\lceil B_n(1-\alpha)\rceil)} \\
&\Leftrightarrow \quad \sum_{b=1}^{B_n} \mathbb{1}_{U^{*b}<U_{n,h}(\mathbb{X}_n)} \geq \lceil B_n(1-\alpha)\rceil \\
&\Leftrightarrow \quad \sum_{b=1}^{B_n} \mathbb{1}_{U^{*b}\geq U_{n,h}(\mathbb{X}_n)} \leq \lfloor \alpha B_n \rfloor \\
&\Leftrightarrow \quad \sum_{b=1}^{B_n} \mathbb{1}_{U^{*b}\geq U_{n,h}(\mathbb{X}_n)} \leq \alpha B_n.
\end{aligned}$$

Un test d'indépendance par permutation

Remarquons que sous l'hypothèse nulle d'indépendance ($\mathcal{H}_0$), la loi de l'échantillon $\mathbb{X}_n$ est invariante sous le groupe de transformations consistant à permuter les secondes coordonnées de $\mathbb{X}_n$. L'hypothèse fondamentale ($\mathcal{A}$) des tests par permutation est donc satisfaite, et une telle approche est donc applicable pour cette problématique.

Commençons par décrire la méthode par permutation considérée dans cette thèse, inspirée de celles de Hoeffding [84] et de van der Vaart et Wellner [173]. Soit Π_n une permutation aléatoire uniformément distribuée sur l'ensemble $\mathfrak{S}_n$ des permutations de $\{1, \ldots, n\}$, indépendante de $\mathbb{X}_n$. L'échantillon permuté est alors défini comme $\mathbb{X}_n^{\Pi_n} = (X_1^{\Pi_n}, \ldots, X_n^{\Pi_n})$ où chaque variable $X_i^{\Pi_n} = (X_i^1, X_{\Pi_n(i)}^2)$ est obtenu en permutant les secondes coordonnées selon Π_n.

Ainsi que pour le test par bootstrap, nous définissons la valeur critique comme le quantile $q^{\star}_{1-\alpha,n}(\mathbb{X}_n)$ d'ordre $(1-\alpha)$ de la loi conditionnelle de la statistique de test permutée $\sqrt{n}U_{n,h}(\mathbb{X}_n^{\Pi_n})$ sachant $\mathbb{X}_n$, et le *test unilatéral à droite par permutation* rejette l'indépendance lorsque

$$\sqrt{n}U_{n,h}(\mathbb{X}_n) > q^{\star}_{1-\alpha,n}(\mathbb{X}_n).$$

La valeur critique, dépendant de l'observation $\mathbb{X}_n$, est également aléatoire, et il est possible de la calculer de manière exacte. Puisque Π_n est uniformément distribué sur $\mathfrak{S}_n$, et indépendant de $\mathbb{X}_n$, la loi conditionnelle de $\sqrt{n}U_{n,h}\big(\mathbb{X}_n^{\Pi_n}\big)$ sachant $\mathbb{X}_n$ est discrète et à valeurs dans $\{\sqrt{n}U_{n,h}(\mathbb{X}_n^{\pi_n})\}_{\pi_n\in\mathfrak{S}_n}$. Soit

$$\sqrt{n}U^{\star(1)}_{n,h}(\mathbb{X}_n) \leq \ldots \leq \sqrt{n}U^{\star(n!)}_{n,h}(\mathbb{X}_n)$$

ces valeurs ré-ordonnées dans l'ordre croissant. Alors, $q^{\star}_{1-\alpha,n}(\mathbb{X}_n) = \sqrt{n}U^{\star(\lceil n!(1-\alpha)\rceil)}_{n,h}(\mathbb{X}_n)$.

Comme pour le bootstrap, des méthodes par Monte Carlo sont appliquées en pratique pour approcher à la fois le quantile et la p-valeur. Cependant, pour des considérations de niveau exact détaillées dans la section 0.2.2, la méthode par permutation diffère légèrement de celle par bootstrap, dans le sens où nous ajoutons la statistique de test calculée sur l'échantillon observé à la famille des statistiques permutées.
Plus précisément, simulons B_n permutations i.i.d. de loi uniforme sur $\mathfrak{S}_n$ et indépendantes de $\mathbb{X}_n$, notées $\Pi_n^1, \ldots, \Pi_n^{B_n}$. Posons

$$U^{\star b} = U_{n,h}\left(\mathbb{X}_n^{\Pi_n^b}\right) \text{ pour tout } 1 \leq b \leq B_n, \quad \text{et} \quad U^{\star B_n+1} = U_{n,h}(\mathbb{X}_n).$$

Considérons la statistique d'ordre $U^{\star(1)} \leq \cdots \leq U^{\star(B_n+1)}$. Alors, le quantile approché est défini par $q^{\star MC}_{1-\alpha,n}(\mathbb{X}_n) = \sqrt{n}U^{\star(\lceil(1-\alpha)(B_n+1)\rceil)}$ et le *test unilatéral à droite par permutation avec approximation par Monte Carlo* rejette l'indépendance lorsque

$$\sqrt{n}U_{n,h}(\mathbb{X}_n) > q^{\star MC}_{1-\alpha,n}(\mathbb{X}_n).$$

La p-valeur correspondante de ce test est définie par

$$\frac{1}{B_n+1}\left(1+\sum_{b=1}^{B_n}\mathbb{1}_{U^{*b}\geq U_{n,h}(\mathbb{X}_n)}\right).$$

Comme pour le test par bootstrap, nous pouvons vérifier que le *test unilatéral à droite par permutation avec approximation par Monte Carlo* est exactement le test rejetant l'indépendance lorsque cette p-valeur est inférieure ou égale à α.

Remarquons la différence entre l'étoile $*$ correspondant à l'approche par bootstrap, et l'étoile $\star$ correspondant à l'approche par permutation. Maintenant que nous avons introduit les différentes procédures de test étudiées dans cette thèse, nous nous interrogeons sur leurs performances.

0.2 Étude des performances

L'étude des performances d'un test passe par l'analyse de ses risques de première et de seconde espèces. Classiquement, lorsque un test est basé sur des approches par bootstrap ou par permutation, seules ses propriétés asymptotiques sont étudiées (voir par exemple [155, 173]). Dans cette thèse, nous avons voulu aller plus loin et étudier les propriétés non-asymptotiques des tests par permutation qui sont connus pour être de niveau exact prescrit.
Le chapitre 1 est consacré à l'étude des performances asymptotiques des tests par bootstrap et par permutation dans le cadre de processus ponctuels, et le chapitre 4 contient l'étude des propriétés non-asymptotiques du test bilatéral par permutation dans le cadre de variables à densité.

0.2.1 Étude asymptotique

Plusieurs manières d'étudier les propriétés asymptotiques d'un test ont été introduites dans la littérature. Celle adoptée dans cette thèse consiste à vérifier séparément que chacun des risques (de première et de seconde espèces) sont contrôlés lorsque la taille de l'échantillon tend vers $+\infty$. Plus précisément, nous vérifions que les tests sont asymptotiquement de taille prescrite, et consistants contre certaines alternatives.
Généralement, lorsque l'on applique une approche par bootstrap (respectivement par permutation) est de démontrer que la loi conditionnelle de la statistique de test bootstrappée (respectivement permutée) sachant l'échantillon initial converge vers la loi asymptotique de la statistique de test sous l'hypothèse nulle. Si ce résultat n'est valable que lorsque l'échantillon initial vérifie l'hypothèse nulle, nous dirons que cette approche *permet de mimer* la loi sous $(\mathcal{H}_0)$. Dans ce cas, puisque ce résultat ne tient que sous l'hypothèse nulle, il ne concerne que la propriété de niveau/taille. Par ailleurs, si ce résultat reste valable sous certaines alternatives, nous dirons que cette approche *permet de reconstruire* la loi sous $(\mathcal{H}_0)$. En plus des propriétés de niveau/taille, cela permet également de déduire la consistance du test sous ces alternatives.

Sur les résultats de van der Vaart et Wellner et de Romano. À titre d'exemple, rappelons les résultats asymptotiques obtenus par van der Vaart et Wellner [173, Section 3.8] pour leurs tests d'indépendance par bootstrap et par permutation, basés sur la statistique de test H_n de type Kolmogorov-Smirnov rappelée dans (0.1.6).

- D'abord, ils démontrent que, si les ensembles $\mathcal{V}^1$ et $\mathcal{V}^2$ sont des classes de Donsker (ce qui signifie de façon informelle qu'il existe un théorème de la limite centrale uniforme sur ces classes, c.f. [173, p.81]), et si $\mathbb{X}_n$ est un échantillon i.i.d. de loi P vérifiant $(\mathcal{H}_0)$, à savoir $P = P^1 \otimes P^2$, alors la loi de $H_n(\mathbb{X}_n)$ converge faiblement vers une loi limite $\mathcal{L}_{P^1 \otimes P^2}$ dépendant uniquement de P^1 et P^2.

- D'une part, pour l'approche par bootstrap, ils prouvent que la loi conditionnelle de la statistique de test bootstrappée sachant $\mathbb{X}_n$ converge faiblement vers la même loi limite $\mathcal{L}_{P^1 \otimes P^2}$, et ce presque sûrement en l'échantillon initial $\mathbb{X}_n$, que P vérifie ou non l'hypothèse nulle. Ainsi, ils montrent que l'approche par bootstrap permet de reconstruire la loi sous $(\mathcal{H}_0)$.

- D'autre part, la question de la capacité à reconstruire la loi sous $(\mathcal{H}_0)$ de l'approche par permutation est laissée ouverte, même s'ils pensent que la statistique de test permutée doit avoir un comportement asymptotique similaire à celui de la statistique de test bootstrappée.

Cette similarité entre le bootstrap et la permutation est montrée par Romano [155, Proposition 3.1] dans le cas particular de fonctions indicatrices sur des ensembles appartenant à une classe de Vapnik-Chervonenkis. Plus précisément, Romano démontre que la norme infinie entre les fonctions de répartition des lois conditionnelles de ses statistiques de test bootstrappée et permutée sachant $\mathbb{X}_n$ converge en probabilité vers zéro, et ce que P vérifie ou non l'hypothèse nulle (puisque les échantillons bootstrappé et permuté sont tous deux forcés de vérifier $(\mathcal{H}_0)$). Pour cela, il commence par montrer, dans [154], des résultats similaires à ceux de van der Vaart et Wellner pour la loi de la statistique de test sous $(\mathcal{H}_0)$ et l'approche par bootstrap. Comme ci-dessus, notons $\mathcal{L}_{P^1\otimes P^2}$ la loi limite commune. Ensuite, basé sur le résultat de Hoeffding [84, Théorème 3.2], il montre que la loi conditionnelle de la statistique de test permutée converge faiblement en probabilité vers la même limite $\mathcal{L}_{P^1\otimes P^2}$, et ce, que l'échantillon initial $\mathbb{X}_n$ vérifie ou non $(\mathcal{H}_0)$. Il démontre donc que les approches par bootstrap et par permutation permettent toutes deux de reconstruire la loi sous $(\mathcal{H}_0)$. Remarquons cependant que, pour démontrer théoriquement la capacité de reconstruction sous l'alternative, il utilise une approche par permutation qui n'est pas exactement celle décrite ici dans le sens où, à notre connaissance, il ne considère pas les permutations ayant des points fixes.

Contributions de cette thèse

Suivant le même raisonnement, nous obtenons les résultats suivants dans le cadre de processus ponctuels. Une fois de plus, seule l'étude des tests unilatéraux à droite est présentée ici par souci de simplicité, mais des résultats similaires pour les tests unilatéraux à gauche et les bilatéraux sont également démontrés dans le chapitre 1. Au lieu de regarder les fonctions de répartition comme le fait Romano, toutes les convergences sont exprimées dans le chapitre 1 en termes de la distance de Wasserstein d'ordre 2 comme cela est généralement le cas lors de l'étude d'approches par bootstrap (voir par exemple [20]). Notamment, rappelons qu'une convergence dans une telle distance implique une convergence faible (voir la proposition A.1.2 en annexe A.1.2).

Résultats de convergence. Dans le contexte où se situe cette thèse, selon le théorème de la limite centrale classique pour les U-statistiques non-dégénérées (voir l'annexe A.3 pour le résultat général, ou la proposition 1.3.5 dans le chapitre 1 pour sa version en distance de Wasserstein d_2), sous certaines hypothèses faibles de moment et de non-dégénérescence, si $P = P^1 \otimes P^2$, alors la loi de la statistique de test $\sqrt{n}U_{n,h}(\mathbb{X}_n)$, notée $\mathcal{L}\left(\sqrt{n}U_{n,h}, P\right)$, converge en distance de Wasserstein vers une loi normale centrée de variance $\sigma^2_{P^1\otimes P^2}$, à savoir

$$d_2\left(\mathcal{L}\left(\sqrt{n}U_{n,h}, P^1\otimes P^2\right), \mathcal{N}\left(0, \sigma^2_{P^1\otimes P^2}\right)\right) \underset{n\to+\infty}{\longrightarrow} 0. \tag{0.2.1}$$

En particulier, la loi de la statistique de test sous l'hypothèse nulle converge vers $\mathcal{N}(0, \sigma^2_{P^1\otimes P^2})$.

Par ailleurs, nous démontrons que les approches par bootstrap et par permutation permettent de reconstruire la loi sous $(\mathcal{H}_0)$. Pour le test par bootstrap, nous prouvons directement (sans utiliser la loi normale asymptotique comme intermédiaire) que la distance entre la loi conditionnelle de la statistique bootstrappée sachant $\mathbb{X}_n$, notée $\mathcal{L}\left(\sqrt{n}U_{n,h}, P_n^1\otimes P_n^2 \middle| \mathbb{X}_n\right)$, et la loi de la statistique de test sous $(\mathcal{H}_0)$ converge vers zéro. Plus précisément, dans le théorème 1.3.1, nous démontrons que

$$d_2\left(\mathcal{L}\left(\sqrt{n}U_{n,h}, P_n^1\otimes P_n^2 \middle| \mathbb{X}_n\right), \mathcal{L}\left(\sqrt{n}U_{n,h}, P^1\otimes P^2\right)\right) \underset{n\to+\infty}{\longrightarrow} 0, \ P\text{-p.s. en } (X_i)_i. \tag{0.2.2}$$

Les arguments dans la preuve sont classiques (voir [20, 41, 118]), basés sur des outils usuels tels que la convergence faible des mesures empiriques dans les espaces séparables [174, Théorème 3] (voir le théorème A.1.4 en annexe A.1.2), et le théorème de représentation de Skorokhod [46, Théorème 11.7.2] (voir le théorème A.1.2 en annexe A.1.2). La plus grande difficulté rencontrée est essentiellement due à la nature de nos variables aléatoires, à savoir des processus ponctuels.

Concernant l'approche par permutation, nous avons dû, pour des raisons techniques, restreindre notre étude au *cas linéaire*, qui correspond au cas où le noyau h de la U-statistique est de la forme h_φ comme dans (0.1.4). Nous avons montré que, sous cette condition, la statistique de test permutée peut être écrite comme une somme permutée, centrée et renormalisée, à savoir

$$\begin{aligned}\sqrt{n}U_{n,h_\varphi}\big(\mathbb{X}_n^{\Pi_n}\big) &= \frac{\sqrt{n}}{n-1}\left(\sum_{i=1}^n \varphi\Big(X_i^1, X_{\Pi_n(i)}^2\Big) - \frac{1}{n}\sum_{i,j}\varphi\big(X_i^1, X_j^2\big)\right)\\ &= \frac{\sqrt{n}}{n-1}\left(\sum_{i=1}^n \varphi\Big(X_i^1, X_{\Pi_n(i)}^2\Big) - \mathbb{E}\left[\sum_{i=1}^n \varphi\Big(X_i^1, X_{\Pi_n(i)}^2\Big)\Big|\mathbb{X}_n\right]\right). \quad (0.2.3)\end{aligned}$$

Une des plus grandes difficultés rencontrées pour cette approche vient du fait que les fonctions φ ne sont plus supposées être de la forme produit, comme c'est le cas dans les tests de Romano [155] ou de van der Vaart et Wellner's [173]. Afin de montrer la validité de cette approche, nous avons démontré un nouveau théorème de la limite centrale combinatoire pour les sommes permutées (voir le théorème 1.4.1 du chapitre 1). Plus précisément, nous avons prouvé que la distance entre la loi conditionnelle de la statistique de test permutée sachant $\mathbb{X}_n$, notée $\mathcal{L}\left(\sqrt{n}U_{n,h_\varphi}, P_n^\star\middle|\mathbb{X}_n\right)$ (où $P_n^\star$ désigne la loi conditionnelle de $\mathbb{X}_n^{\Pi_n}$ sachant $\mathbb{X}_n$), et la loi asymptotique de la statistique de test sous $(\mathcal{H}_0)$ obtenue dans (0.2.1) converge vers zéro, à savoir

$$d_2\left(\mathcal{L}\left(\sqrt{n}U_{n,h_\varphi}, P_n^\star\middle|\mathbb{X}_n\right), \mathcal{N}\left(0, \sigma^2_{P^1\otimes P^2}\right)\right) \underset{n\to+\infty}{\longrightarrow} 0, \text{ en probabilité.} \quad (0.2.4)$$

Remarquons que de nombreux théorèmes de la limite centrale combinatoires ont été montrés dans la littérature. Le premier résultat, obtenu par Wald et Wolfowitz [178] pour des sommes de terme déterministes de la forme produit, permutées aléatoirement, telles que $\sum_{i=1}^n b_i \times c_{\Pi_n(i)}$, a été largement généralisé : par exemple à des termes non-nécessairement de la forme produit par Hoeffding [83], sous des hypothèse de type Lindeberg par Motoo [127], à des termes aléatoires par Dwass [49], et plus récemment, à des U-statistiques pondérées par Shapiro et Hubert [165]. Plus de détails peuvent être trouvés dans l'introduction du chapitre 3. La nouveauté principale de ce résultat réside dans le fait qu'il n'a pas besoin de l'hypothèse habituelle et restrictive d'échangeabilité sur les variables aléatoires permutées, et donc, reste vrai sous n'importe quelle alternative. Notamment, il implique que l'approche par permutation permet de reconstruire (et pas seulement de mimer) la loi de la statistique de test sous $(\mathcal{H}_0)$, et ainsi, répond partiellement (nous ne considérons pas de supremum dans notre statistique de test) à la question ouverte laissée par van der Vaart et Wellner dans [173]. C'est l'un des résultats les plus novateurs parmi ceux présentés ici, et son étendue va au delà de la simple généralisation au cas de processus ponctuels, puisque sa démonstration reste valable pour n'importe quelles variables aléatoires i.i.d. à valeurs dans un espace séparable.

Propriétés asymptotiques des tests. Finalement, nous déduisons de ces résultats de convergence que les tests par bootstrap et par permutation sont asymptotiquement de taille

voulue, et consistant contre n'importe quelle alternative raisonnable P sous laquelle l'espérance de la statistique de test, précisément $\int_{\mathcal{X}^2\times\mathcal{X}^2} h(x,y)dP(x)dP(y)$, est strictement positive (nous rappelons qu'elle est égale à zéro sous l'hypothèse nulle à cause de l'hypothèse de recentrage). Les preuves sont classiques, et leur heuristique est décrite pour le test par bootstrap (respectivement par permutation) ci-dessous.
D'une part, la taille asymptotique est obtenue comme suit. Supposons que P vérifie $(\mathcal{H}_0)$.

- Premièrement, grâce à (0.2.1), nous obtenons que la statistique de test $\sqrt{n}U_{n,h}(\mathbb{X}_n)$ converge en loi vers une variable aléatoire de loi $\mathcal{N}\left(0,\sigma^2_{P^1\otimes P^2}\right)$, disons $Z_{P^1\otimes P^2}$.
- Ensuite, la loi asymptotique étant continue, en combinant (0.2.1) et (0.2.2) (respectivement grâce à (0.2.4)), nous déduisons que la valeur critique $q^*_{1-\alpha,n}(\mathbb{X}_n)$ (respectivement $q^\star_{1-\alpha,n}(\mathbb{X}_n)$), en tant que quantile conditionnel, converge presque sûrement (respectivement en probabilité) vers le quantile d'ordre $(1-\alpha)$ de la loi limite $\mathcal{N}\left(0,\sigma^2_{P^1\otimes P^2}\right)$, noté $\Phi^{-1}_{0,\sigma^2_{P^1\otimes P^1}}(1-\alpha)$.
- Par conséquence, en combinant ces résultats au lemme de Slutsky (rappelé dans la proposition A.1.1 en annexe A.1.2), la probabilité que le test par bootstrap rejette à tort $(\mathcal{H}_0)$ vérifie

$$\mathbb{P}\big(\sqrt{n}U_{n,h}(\mathbb{X}_n) > q^*_{1-\alpha,n}(\mathbb{X}_n)\big) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}\left(Z_{P^1\otimes P^2} > \Phi^{-1}_{0,\sigma^2_{P^1\otimes P^1}}(1-\alpha)\right) = \alpha.$$

  Le test par permutation vérifie un résultat similaire, précisément

$$\mathbb{P}\big(\sqrt{n}U_{n,h}(\mathbb{X}_n) > q^\star_{1-\alpha,n}(\mathbb{X}_n)\big) \underset{n\to+\infty}{\longrightarrow} \alpha.$$

Ceci étant vrai pour n'importe quelle loi sous-jacente P vérifiant $(\mathcal{H}_0)$, le test par bootstrap (respectivement par permutation) est bien de taille asymptotique voulue α.
D'autre part, la preuve la consistance des tests repose également la loi des grands nombres pour les U-statistiques (rappelée dans le théorème A.3.1 en annexe A.3.2) stipulant que

$$U_{n,h}(\mathbb{X}_n) \underset{n\to+\infty}{\longrightarrow} \int_{\mathcal{X}^2\times\mathcal{X}^2} h(x,y)dP(x)dP(y), \ P\text{-p.s. en } (X_i)_i. \tag{0.2.5}$$

Par ailleurs, (0.2.2) (respectivement (0.2.4)) étant valable aussi sous n'importe quelle alternative, la valeur critique $q^*_{1-\alpha,n}(\mathbb{X}_n)$ (respectivement $q^\star_{1-\alpha,n}(\mathbb{X}_n)$) converge toujours vers le quantile gaussien $\Phi^{-1}_{0,\sigma^2_{P^1\otimes P^1}}(1-\alpha)$. Ainsi, sous toute alternative raisonnable P telle que la limite dans (0.2.5) est strictement positive, nous obtenons

$$\begin{aligned}\mathbb{P}\big(\sqrt{n}U_{n,h}(\mathbb{X}_n) > q^*_{1-\alpha,n}(\mathbb{X}_n)\big) &= \mathbb{P}\left(U_{n,h}(\mathbb{X}_n) > \frac{q^*_{1-\alpha,n}(\mathbb{X}_n)}{\sqrt{n}}\right)\\ &\underset{n\to+\infty}{\longrightarrow} \mathbb{P}\left(\int_{\mathcal{X}^2\times\mathcal{X}^2} h(x,y)dP(x)dP(y) > 0\right) = 1.\end{aligned}$$

Un résultat similaire vaut également pour le test par permutation en remplaçant $q^*_{1-\alpha,n}(\mathbb{X}_n)$ par $q^\star_{1-\alpha,n}(\mathbb{X}_n)$. Ce résultat restant vrai sous n'importe quelle alternative P vérifiant $\int_{\mathcal{X}^2\times\mathcal{X}^2} h(x,y)dP(x)dP(y) > 0$, le test par bootstrap (respectivement par permutation) est donc consistant contre chacune de ces alternatives. Remarquons que la capacité à reconstruire la loi sous $(\mathcal{H}_0)$ est une propriété plus forte que la simple consistance du test, qui a déjà été démontrée dans certains cas, notamment par Hoeffding [84] pour les tests par permutation.

Nous avons également démontré que les tests par bootstrap et par permutation avec approximation par Monte Carlo sont de taille asymptotique voulue et consistant contre les mêmes alternatives. Intuitivement, cela découle de la convergence des quantiles approchés vers la même limite que les quantiles exacts, à savoir le quantile de la loi asymptotique sous $(\mathcal{H}_0)$.

0.2.2 Étude non-asymptotique

Ces résultats asymptotiques justifient dans un premier temps l'application de telles approches pour tester l'indépendance. Cependant, comme nous le mentionnons au début de cette introduction, dans de nombreux cas, et en particulier lors d'applications en biologie comme dans le chapitre 2, peu de données sont exploitables, et des résultats purement asymptotiques peuvent être insuffisants pour pleinement justifier d'un point de vue théorique l'application de telles procédures. Pour cette raison, nous avons également étudié les propriétés non-asymptotiques des tests par permutation. Dans cette thèse, ces propriétés sont étudiées en termes de niveau exact et de vitesses de séparation uniforme introduites en section 0.1.2.

Comme précisé ci-dessus, un des avantages principaux de l'approche par permutation est que, lorsqu'elle est appliquée correctement, elle permet de construire des procédures de test dont le risque de première espèce est contrôlé (non-asymptotiquement) par un niveau prescrit. Ce résultat est classique, mais pour que ce manuscrit soit auto-contenu, nous allons montrer que le test unilatéral à droite par permutation introduit dans la section 0.1.3 est en effet de niveau exact prescrit α. Soit P vérifiant $(\mathcal{H}_0)$ et $\mathbb{X}_n$ un échantillon de variables aléatoires i.i.d. de loi P. Rappelons d'abord que l'échangeabilité des variables sous $(\mathcal{H}_0)$ signifie ici que pour toute permutation π_n dans $\mathfrak{S}_n$, $\mathbb{X}_n^{\pi_n}$ et $\mathbb{X}_n$ sont de même loi. Ainsi

$$\mathbb{P}\big(\sqrt{n}U_{n,h}(\mathbb{X}_n) > q^\star_{1-\alpha,n}(\mathbb{X}_n)\big) = \frac{1}{n!}\sum_{\pi_n\in\mathfrak{S}_n} \mathbb{P}\big(\sqrt{n}U_{n,h}(\mathbb{X}_n^{\pi_n}) > q^\star_{1-\alpha,n}(\mathbb{X}_n^{\pi_n})\big)\,.$$

De plus, les quantiles conditionnels étant obtenus à partir de la statistique d'ordre considérant toutes les permutations de $\mathbb{X}_n$, ils sont donc invariants par permutation, c'est-à-dire que pour tout π_n dans $\mathfrak{S}_n$, $q^\star_{1-\alpha,n}(\mathbb{X}_n^{\pi_n}) = q^\star_{1-\alpha,n}(\mathbb{X}_n)$.
Ainsi, nous obtenons que

$$\begin{aligned}
\mathbb{P}\big(\sqrt{n}U_{n,h}(\mathbb{X}_n) > q^\star_{1-\alpha,n}(\mathbb{X}_n)\big) &= \frac{1}{n!}\sum_{\pi_n\in\mathfrak{S}_n} \mathbb{P}\big(\sqrt{n}U_{n,h}(\mathbb{X}_n^{\pi_n}) > q^\star_{1-\alpha,n}(\mathbb{X}_n)\big)\\
&= \sum_{\pi_n\in\mathfrak{S}_n} \mathbb{P}\big(\sqrt{n}U_{n,h}\big(\mathbb{X}_n^{\Pi_n}\big) > q^\star_{1-\alpha,n}(\mathbb{X}_n)\big|\Pi_n=\pi_n\big)\,\mathbb{P}(\Pi_n=\pi_n)\\
&= \mathbb{P}\big(\sqrt{n}U_{n,h}\big(\mathbb{X}_n^{\Pi_n}\big) > q^\star_{1-\alpha,n}(\mathbb{X}_n)\big)\\
&= \mathbb{E}\left[\mathbb{P}\big(\sqrt{n}U_{n,h}\big(\mathbb{X}_n^{\Pi_n}\big) > q^\star_{1-\alpha,n}(\mathbb{X}_n)\big|\mathbb{X}_n\big)\right]\\
&\leq \alpha,
\end{aligned}$$

par définition du quantile conditionnel. Ceci étant vrai quelle que soit la loi sous-jacent P vérifiant l'hypothèse nulle, le test est bien de niveau exact α.
En général, aucun résultat équivalent n'existe pour l'approche par bootstrap. Par conséquent, lorsque les deux approches sont possibles, il vaut mieux privilégier celle par permutation (voir par exemple [51]). Pour cette raison, nous étudions uniquement le test par permutation d'un point de vue non-asymptotique.

Par ailleurs, il résulte directement du lemme de Romano et Wolf [156, Lemme 1] que, grâce à l'échangeabilité des variables aléatoires sous l'hypothèse nulle, le test par permutation avec l'approximation de Monte Carlo, tel que celui introduit dans la section 0.1.3, est également de niveau exact α. La preuve de ce lemme est donnée par Blanchard et al. dans [7].

Il reste alors le contrôle non-asymptotique du risque de seconde espèce. Les vitesses (asymptotiques) de test minimax des tests d'indépendance entre d variables aléatoires réelles ont été étudiées dans la littérature, d'abord par Ingster [94], suivi de Yodé [180] sur des classes de Hölder de paramètre de régularité ν par rapport à différentes distance. En particulier, ils ont montré qu'elles sont égales à

$$n^{\frac{-2\nu}{4\nu+d}} \text{ par rapport à la distance } \mathbb{L}_2, \quad \text{et} \quad \left(\frac{n}{\ln(n)}\right)^{\frac{-\nu}{2\nu+d}} \text{ par rapport à la distance } \mathbb{L}_\infty.$$

Ensuite, Yodé propose dans [181] une procédure de test adaptative au sens du minimax, basée sur les U-statistiques, atteignant la vitesse de test minimax par rapport à la distance $\mathbb{L}_2$ à un facteur logarithmique près, ce qui est un prix usuel et raisonnable à payer pour l'adaptativité. Cependant, ces approches sont purement asymptotiques, et ne contrôlent pas le niveau pour des tailles d'échantillon n petites (il est seulement démontré que le risque de première espèce est contrôlé par une suite tendant vers α lorsque n tend vers l'infini). Néanmoins, ces résultats sont d'importance capitale ici étant donné qu'ils fournissent les vitesses de test minimax, qui devraient être équivalentes aux *vitesses de séparation minimax*, définies dans la section 0.1.2, qui nous intéressent dans cette thèse.
De nombreuses procédures de test adaptatives au sens du minimax ont été développées dernièrement dans de nombreux contextes, basées sur des méthodes de sélection de modèles (voir par exemple [14, 55]), ou de seuillage (voir par exemple [167]). Dans cette thèse, nous considérons une méthode de seuillage par ondelettes détaillée ci-dessous.

Contributions de cette thèse

Comme expliqué ci-dessus, seuls les tests par permutation sont étudiés d'un point de vue non-asymptotique.
Afin de ramener notre étude à des contextes plus classique dans lesquels les vitesses minimax sont connues, l'analyse des vitesses de séparation uniforme effectuée dans le chapitre 4 est faite dans un cadre de variables aléatoires à densité. Dans ce cas, l'hypothèse nulle équivaut à l'égalité de la fonction de densité f, et le produit de ses marginales $f^1 \otimes f^2$. Dans un but d'adaptativité, nous construisons une procédure de tests multiples basée sur l'agrégation de plusieurs tests simples inspirée des méthodes de seuillage par ondelettes considérées par Fromont et al. dans [57] ou par Sansonnet et Tuleau-Malot dans [160]. Lors de l'utilisation de bases d'ondelettes, il est naturel de considérer des espaces de régularité tels que les espaces de Besov (voir par exemple [13, 32, 55, 57, 60, 98, 99, 116, 160, 167]) ou des version faibles de ces derniers (voir par exemple [57, 58, 160]), et la distance $\mathbb{L}_2$ étant donné qu'ils peuvent tous être définis de manière simple en fonction des coefficients d'ondelettes. Plus de détails sont disponibles dans l'annexe A.4.
Soient deux niveaux d'erreurs α et β dans $]0,1[$ et soit $\mathbb{X}_n$ un échantillon de variables aléatoires i.i.d. de loi P.

La décomposition en ondelettes. Soit $\{\varphi_\lambda\}_{\lambda\in\Lambda}$ une base d'ondelettes de $\mathbb{L}_2([0,1]^2)$. Dans cette thèse, nous considérons la base de Haar (voir la section 4.2.1 du chapitre 4, ou

l'annexe A.4.1) par simplicité, mais les résultats peuvent être généralisés à des bases plus régulières (voir [39] ou [126] par exemple). Pour tout indice λ dans Λ, nous introduisons le coefficient de la différence entre la fonction de densité f et le produit de ses marginales $f_1 \otimes f_2$, défini comme $\beta_\lambda = \int_{[0,1]^2} [f(x) - f_1 \otimes f_2(x)] \varphi_\lambda(x) dx$. Remarquons que sous l'hypothèse nulle, chaque coefficient β_λ est égal à zéro.

Les tests individuels de coefficient. Pour chaque élément de la base, nous construisons un test individuel bilatéral basé sur la même approche par permutation que dans la section 0.1.3. Plus précisément, fixons un indice λ dans Λ et définissons la nouvelle statistique de test par $|T_\lambda(\mathbb{X}_n)|$ où $T_\lambda(\mathbb{X}_n) = U_{n,h_{\varphi_\lambda}}(\mathbb{X}_n)$ (voir (0.1.3) et (0.1.4)). Nous oublions le terme de renormalisation $\sqrt{n}$ pour l'étude d'un point de vue non-asymptotique. En particulier, $T_\lambda(\mathbb{X}_n)$ est un estimateur sans biais du coefficient β_λ. Ainsi, le *test de coefficient bilatéral par permutation* associé à λ rejette l'indépendance lorsque

$$|T_\lambda(\mathbb{X}_n)| > q_{\lambda,1-\alpha}(\mathbb{X}_n),$$

$q_{\lambda,1-\alpha}(\mathbb{X}_n)$ désignant le quantile d'ordre $(1-\alpha)$ de la loi conditionnelle de la statistique de test permutée $\left|T_\lambda(X_n^{\Pi_n})\right|$ sachant $\mathbb{X}_n$ où Π_n est une permutation aléatoire uniforme de $\{1, \dots, n\}$ indépendante de $\mathbb{X}_n$.

Par construction, le test est bien de niveau exact prescrit α. Rappelons que notre objectif est de majorer la vitesse de séparation du test par la vitesse de séparation minimax sur certains espaces de régularité par rapport à la distance $\mathbb{L}_2$ à constante près. Dans un premier temps, nous obtenons une condition sur l'alternative f garantissant un contrôle du risque de seconde espèce par la valeur prescrite β. Plus précisément, le coefficient $\beta_\lambda = \beta_\lambda(f)$ étant égal à zéro sous $(\mathcal{H}_0)$, plus il est grand, plus l'alternative f est loin de $(\mathcal{H}_0)$, et donc plus le test est capable de la détecter. Ainsi, nous trouvons un seuil s, dépendant de n, α, β, φ_λ et $\|f\|_\infty$ tel que si $|\beta_\lambda| \geq s$, alors la probabilité sous l'alternative f d'accepter à tort $(\mathcal{H}_0)$ est contrôlée par β. Ce résultat est basé sur une inégalité de concentration pour les sommes permutées présentée dans la section 0.2.3.

La procédure de tests agrégés. Afin d'éviter le choix délicat du coefficient λ, et pour pouvoir détecter des formes plus générales de dépendance, l'idée est d'agréger plusieurs tests de coefficient jusqu'à une certaine échelle $\tilde{J}$ dans la décomposition en ondelettes. De même que dans la théorie des tests multiples, l'agrégation de tests nécessite une correction des niveaux de chaque test individuel. Ici, nous considérons la même correction que dans [14, 57, 58], qui consiste en la correction la plus fine des niveaux individuels permettant d'appliquer simultanément chaque test de coefficient avec un niveau global α (voir la section 4.2.3 dans le chapitre 4). Ainsi, le *test bilatéral agrégé par permutation* Δ_α rejette l'hypothèse nulle si au moins l'un des tests (individuels corrigés) de coefficient bilatéraux par permutation la rejette.

Par construction de la correction des niveaux individuels, le test agrégé est de niveau exact α. Ensuite, pour l'étude des vitesses de séparation uniforme, nous considérons un espace de régularité $\mathcal{BW}_{\delta,\gamma,\infty}(R, R', R'')$ défini proprement dans le chapitre 4, et étant l'intersection d'un espace de Besov $\mathcal{B}^\delta_{2,\infty}(R)$ de paramètre de régularité $\delta > 0$ et de rayon $R > 0$, d'un espace de Besov faible $\mathcal{W}_\gamma(R')$ de paramètre de régularité $\gamma > 0$ et de rayon $R' > 0$ (voir l'annexe A.4.2), et de la boule $\mathbb{L}_\infty$ de rayon $R'' > 0$. Finalement, nous obtenons que si l'échelle maximale $\tilde{J}$

est égale à $\lfloor \log_2 \sqrt{n/\ln(n)} \rfloor$, et si $\delta \geq \gamma/(\gamma+1)$, alors, pour n suffisamment grand,

$$\rho\left(\Delta_\alpha, \mathcal{BW}_{\delta,\gamma,\infty}(R,R',R''),\beta\right) \leq C(\alpha,\beta,\delta,\gamma,R,R',R'')\left(\frac{n}{\ln(n)}\right)^{\frac{-\gamma}{2\gamma+2}},$$

où $C(\alpha,\beta,\delta,\gamma,R,R',R'')$ désigne une constante strictement positive. Même si les vitesses de séparation minimax sur de tels espaces de régularité par rapport à la distance $\mathbb{L}_2$ n'ont pas encore été déterminées, cette vitesse de séparation uniforme semble être optimale au sens du minimax au vue de la littérature, et plus précisément, des résultats de Ingster [94] et de Yodé [180, 181]. De plus, la procédure de test ne dépendant pas des paramètres de régularité δ et γ, cette procédure semble être adaptative au sens du minimax sur de tels espaces de régularité. Une discussion détaillée est faite dans l'introduction du chapitre 4.

0.2.3 Inégalités de concentration : un outil performant

Les inégalités de concentration sont des outils performants pour l'étude des vitesses de séparation uniforme, comme par exemple dans [58, 160]. Elles permettent de contrôler la probabilité qu'une variable aléatoire $Z_n = \zeta(Y_1,\ldots,Y_n)$, fonction d'une ou plusieurs variables aléatoires $Y_1,\ldots,Y_n$, s'écarte de sa médiane ou de sa moyenne par un nombre réel prescrit. En particulier, dans le cadre de cette thèse, elles fournissent des majorations précises des quantiles conditionnels de la statistique de test permutée.

La littérature sur les inégalités de concentration est colossale (voir les livres de Ledoux [114], de Massart [123], ou celui très récent de Boucheron, Lugosi, et Massart [26] pour des états de l'art très complets). Initialement, elles sont apparues pour des sommes de variables aléatoires indépendantes, telles que, par exemple l'inégalité de Hoeffding [86], celle de Bennett [18] et celle de Bernstein [19]. D'intérêt majeur ici, rappelons cette dernière, énoncée par exemple par Massart dans [123, Proposition 2.9 et Corolaire 2.10].

Theorem 0.2.1 (Inégalité de Bernstein)**.** *Soient $X_1,\ldots,X_n$ des variables aléatoires réelles indépendantes. Supposons qu'il existe deux réels strictement positifs v et c tels que*

$$\sum_{i=1}^n \mathbb{E}\left[X_i^2\right] \leq v \quad \text{et} \quad \forall k \geq 3,\ \sum_{i=1}^n \mathbb{E}\left[(X_i)_+^k\right] \leq \frac{k!}{2} v c^{k-2},$$

où $(\cdot)_+ = \max\{\cdot, 0\}$ désigne la partie positive.

Soit $S = \sum_{i=1}^n (X_i - \mathbb{E}[X_i])$. Alors pour tout $x > 0$,

$$\mathbb{P}\left(S \geq \sqrt{2vx} + cx\right) \leq \exp(-x). \tag{0.2.6}$$

De plus, pour tout $t > 0$,

$$\mathbb{P}(S \geq t) \leq \exp\left(-\frac{t^2}{2(v+ct)}\right). \tag{0.2.7}$$

Remarquons que les deux formes de l'inégalité de Bernstein apparaissent dans la littérature. Cependant, à cause de sa forme, (0.2.6) est souvent préférée en statistique pour le contrôle des quantiles, même si (0.2.7) est plus classique.

Nous dirons qu'une variables aléatoire $Z_n = \zeta(Y_1 \dots, Y_n)$ vérifie une *inégalité de type Bernsetin* si il existe des constantes universelles c_0 et c_1 et des réels strictement positifs c et v dépendant respectivement de la loi des Y_i et précisément des $\mathbb{E}\left[Y_i^2\right]$, tels que pour tout $x > 0$,

$$\mathbb{P}\Big(Z_n \geq \sqrt{2vx} + cx\Big) \leq c_0 \exp\left(-c_1 x\right),$$

ou, plus classiquement, si pour tout $t > 0$,

$$\mathbb{P}(Z_n \geq t) \leq c_0 \exp\left(-\frac{c_1 t^2}{2(v + ct)}\right).$$

Dans cette thèse, nous nous intéressons à la concentration des sommes permutées, qui sont des sommes de variables dépendantes (précisément $Y_i = a_{i,\Pi_n(i)}$, où les $a_{i,j}$ sont des nombres réels, et Π_n est une permutation uniforme de $\{1, \dots, n\}$). Différentes approches ont été introduites afin d'étudier la concentration de fonctions de variables aléatoires dépendantes. Une première approche, comme expliqué dans [168], consiste à décomposer Z_n comme la somme d'une suite de différences de martingales et d'appliquer des inégalités adaptées aux martingales. Par exemple, Delyon a obtenu des inégalités exponentielles pour des sommes de variables faiblement dépendantes dans [42], en se basant sur la théorie des martingales.
Une autre approche possible est basée sur la méthode de Stein qui consiste à approcher des lois compliquées (telles que la loi d'une somme de variables dépendantes) par une loi plus simple et identifiables grâce à l'introduction d'un opérateur caractérisant. Basé sur une telle approche, Chatterjee [33, Proposition 1.1] obtient une inégalité de concentration pour des sommes permutées, rappelée dans le théorème 3.1.3 du chapitre 3, et stipulant que, pour tout $t > 0$,

$$\mathbb{P}(|Z_n - \mathbb{E}[Z_n]| \geq t) \leq 2 \exp\left(-\frac{t^2}{4\mathbb{E}[Z_n] + 2t}\right). \tag{0.2.8}$$

Au lieu d'utiliser de telles approches, nous nous intéressons notamment aux inégalités de concentration pour les permutations aléatoires, et plus particulièrement à l'inégalité fondamentale de Talagrand [168, Théorème 5.1] détaillée ci-dessous, et rappelée dans le théorème 3.1.1 du chapitre 3.
Premièrement, Talagrand introduit une notion de distance entre une permutation π_n dans $\mathfrak{S}_n$ et un sous-ensemble A de $\mathfrak{S}_n$. Pour cela, il se ramène à un espace d'étude plus simple, à savoir $[0,1]^n$, en posant

$$U_A(\pi_n) = \left\{ s \in \{0,1\}^n \ ; \ \exists \tau \in A \text{ vérifiant } \forall 1 \leq i \leq n, \ s_i = 0 \implies \tau(i) = \pi_n(i) \right\}.$$

En particulier, la permutation π_n appartient à A si et seulement si le vecteur nul appartient à $U_A(\pi_n)$. Puis, il définit la distance entre π_n et A comme étant le carré de la distance classique ℓ_2 entre le vecteur nul et l'enveloppe convexe de $U_A(\pi_n)$ dans $[0,1]^n$, notée $V_A(\pi_n)$, ou plus précisément

$$f(A, \pi_n) = \min\left\{ \sum_{i=1}^{n} v_i^2 \ ; \ v = (v_i)_{1 \leq i \leq n} \in V_A(\pi_n) \right\}.$$

En particulier, il est important de garder à l'esprit que la permutation π_n appartient à A si et seulement si $f(A, \pi_n) = 0$.
Ensuite, l'inégalité de Talagrand pour les permutation aléatoires affirme que, si P_n désigne la loi uniforme sur $\mathfrak{S}_n$, alors

$$\int_{\mathfrak{S}_n} \exp\left(\frac{1}{16} f(A, \pi_n)\right) dP_n(\pi_n) \leq \frac{1}{P_n(A)}.$$

Les inégalités de concentration sont alors déduite grâce à l'inégalité de Markov, et en particulier, pour tout $t > 0$,

$$P_n\left(\pi_n \ ; \ f(A, \pi_n) \geq t^2\right) \leq \frac{\exp\left(-t^2/16\right)}{P_n(A)}.$$

Basés sur cette inégalité, de nombreuses inégalités de concentration ont été obtenues par exemple par McDiarmid [124] sous des conditions de type Lipschitz, ou plus récemment par Adamczak et al. [2] pour des fonctions convexes et lipschitziennes (voir le théorème 3.1.2 dans le chapitre 3).

L'inégalité de Talagrand pour les permutations uniformes mène généralement à des inégalités autour de la médiane, alors que nous souhaitons obtenir des inégalités de concentration autour de la moyenne (car dans le *cas linéaire*, la statistique de test s'écrit comme une somme permutée recentrée, c.f. (0.2.3)). Or, Ledoux [114] montre que, dans le cas d'inégalités de concentration exponentielles, la moyenne et la médiane sont équivalentes. Ainsi, les inégalités autour de la moyenne sont les mêmes, à constante près, que celles autour de la médiane.

Contribution de cette thèse

Soit $\{a_{i,j}\}_{1\leq i,j\leq n}$ une collection de nombres réels, et notons pour toute permutation π_n dans $\mathfrak{S}_n$, $Z_n(\pi_n) = \sum_{i=1}^n a_{i,\pi_n(i)}$ la somme permutée correspondante. Nous souhaitons obtenir une inégalité de concentration exponentielle pour la somme permutée aléatoirement $Z_n(\Pi_n)$. Les résultats de McDiarmid [124] ou de Adamczak et al. [2] ne semblent pas adaptés dans notre cas car les conditions de type Lipschitz sont trop restrictives. Plus particulièrement, dans l'étude des vitesses de séparation uniforme du chapitre 4, ces conditions ne sont pas satisfaites. Ainsi, comme expliqué dans le chapitre 3, l'idée est alors d'exploiter la forme avantageuse de somme. Basés sur l'inégalité de Talagrand pour les permutations uniformes, nous améliorons le résultat de Chatterjee [33] en obtenant un terme de variance à la place du terme d'espérance dans l'exponentielle dans (0.2.8).

Nous suivons le raisonnement de Adamczak et al. [2] et appliquons l'inégalité de Talagrand pour les permutations aléatoires. En particulier, nous obtenons dans un premier temps une inégalité grossière pour $\sqrt{Z_n(\Pi_n)}$ autour de sa médiane dans le cas positif ($a_{i,j} \geq 0$ pour tout $1 \leq i, j \leq n$) qui est insuffisante. Dans un deuxième temps, l'idée est alors d'utiliser cette première inégalité dans la preuve afin d'affiner le résultat. Nous passons de la médiane à la moyenne grâce à l'astuce de Ledoux [114]. Cette étape mène malheureusement à des constantes très grandes. Finalement, nous obtenons une *inégalité de type Bernstein* dans le cas général (non-nécessairement positif), qui stipule que pour tout $x > 0$,

$$\mathbb{P}\left(|Z_n - \mathbb{E}[Z_n]| \geq 2\sqrt{2\left(\frac{1}{n}\sum_{i,j=1}^n a_{i,j}^2\right)x} + 2\max_{1\leq i,j\leq n}|a_{i,j}|\, x\right) \leq c_0 \exp\left(-c_1 x\right),$$

où c_0 et c_1 sont des constantes universelles strictement positives.

0.3 Application à la détection de synchronisations en neurosciences

0.3.1 Contexte biologique et motivation

Comprendre le fonctionnement du cerveau est un immense défi. L'information neuronale est transmise à travers notre cerveau en tant que signaux électriques et chimiques, grâce à des cellules nerveuses, à savoir les neurones. Schématiquement, les neurones communiquent entre eux de la manière suivante. Un neurone reçoit des signaux d'entrée provenant de plusieurs neurones au niveau de ses dendrites. Ces signaux augmentent le potentiel électrique de sa membrane au niveau du soma. Si ce potentiel dépasse un certain seuil, dit d'excitation, il augmente alors rapidement puis retombe. Cette dépolarisation brève et stéréotypée, appelée potentiel d'action, se propage le long de l'axone, et l'on parle alors de décharge du neurone. Lorsque les potentiels d'action atteignent les synapses, celles-ci envoient de nouveaux signaux (chimiques) aux dendrites d'un ou plusieurs neurones dits post-synaptiques.

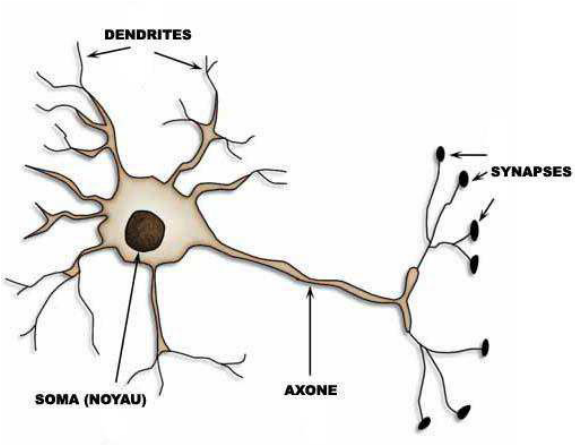


Figure 1 – Schéma d'un neurone.

Il est aujourd'hui communément admis que les potentiels d'actions constituent une des composantes majeurs de l'activité cérébrale (voir [166]). Nous appellons la suite des temps d'occurrences des potentiels d'action d'un neurone un *train de spikes*, les *spikes* étant les temps d'occurrence eux-mêmes. Ces trains de spikes sont généralement modélisés par des processus ponctuels (voir la section 0.3.2 pour plus de détails), chaque point représentant un temps de spike. Le nombre moyen de spikes par unité de temps est appelé *taux de décharge*. Les potentiels d'actions sont déclenchés de deux façons possibles ; soit les neurones pré-synaptiques augmentent (indépendamment) leur taux de décharge instantané, ce qui augmente alors la fréquence des signaux d'entrée (voir [15]), soit ils coordonnent leur activité, de manière à ce que les signaux d'entrée arrivent presque au même moment (voir [79]). Ce deuxième phénomène, appelé *synchronisation*, nécessite moins de signaux d'entrée, utilise moins d'énergie, et augmente la vitesse de transmission de l'information neuronale. Il est aujourd'hui communément admit que ces synchronisations jouent un rôle important dans l'activité neuronale (voir [171]).

Une première étape pour comprendre le code neuronal est de pouvoir détecter de telles synchronisations entre deux neurones ou plus. D'une part, ceci nécessite le développement de techniques expérimentales permettant d'enregistrer simultanément l'activité de plusieurs neu-

rones. Ceci est possible de nos jours grâce à des techniques telles que la combinaison de multi-microélectrodes enregistrant le potentiel extra-cellulaire à proximité des neurones étudiés, et d'un algorithme appelé *spike sorting* permettant d'isoler les signaux des neurones concernés. D'autre part, de nombreuses méthodes statistiques analysant ces synchronisations ont été développées dans la littérature en neurosciences (voir, par exemple, les états de l'art de Harrison et al. [77] ou de Grün [68]). Ces méthodes sont généralement basées sur la notion de *coïncidence* entre deux neurones ou plus, qui apparaît lorsque les neurones étudiés déchargent presque au même moment. Étant donné que, lors d'une synchronisation, les neurones coordonnent leur activité, leur trains de spikes deviennent dépendants. L'idée générale est donc de comparer le nombre de coïncidences observé à celui moyen sous l'hypothèse d'indépendance ; s'il est plus grand que prévu, la corrélation est dite *excitatrice*, et au contraire, s'il est plus petit que prévu, la corrélation est dite *inhibitrice*.

Un outil populaire pour illustrer de tels phénomènes de synchronisation entre deux neurones est le *cross-corrélogramme* (voir [133] ou [77] pour une description détaillée de cette méthode). En quelques mots, cela représente l'histogramme des écarts entre les temps de spikes des deux neurones. Un pic dans le cross-correlogramme représente une corrélation entre les deux neurones étant donné que cela montre une tendance à décharger à un écart privilégié. Une synchronisation est généralement identifiée par de tels pics pour des écarts très proches de zéro. Cependant, il se peut que ces méthodes ne soient pas suffisamment précises, étant principalement visuelles, et appartenant donc plutôt au domaine de la statistique descriptive. En particulier, en tant que tels, ils ne permettent pas d'évaluer la significativité statistique du résultat obtenu. Harrison et al. ont developpé dans [77] une méthode basée sur du bootstrap, sous des hypothèse de modèles de type Poisson homogènes, construisant des régions de confiance pour les cross-corrélogrammes (l'indépendance est rejetée si le cross-corrélogramme sort de cette région). Cependant, l'hypothèse de modèle poissonniens est discutable et même rejetée par certains jeux de données (comme par exemple dans [10, 53, 147]). D'autre méthodes ont également été introduites.

Une des méthodes les plus connues et appliquées ces dernières années est la célèbre méthode *Unitary Events* (UE), introduite par Grün et ses co-auteurs à la fin des années 90 (voir [67, 71]). Cette méthode permet de détecter des groupes fonctionnels de deux neurones ou plus qui sont corrélés (de manière excitatrice ou inhibitrice). Plus précisément, lors de l'observation de L neurones différents, on peut souhaiter détecter des formes particulières de *motifs* (par exemple, tous les neurones déchargent ensemble, ce qui correspond au motif $\{1, \dots, L\}$, et qui est notre cas avec $L = 2$, ou encore, les neurones 1, 3 et 4 déchargent ensemble, correspondant au motif $\{1, 3, 4\}$). Il existe 2^L motifs possibles, et chacun est étudié séparément. Fixons à présent un motif. Un des avantages principaux de la méthode UE initiale est que, étant donnés n observations de L trains de spikes enregistrés simultanément, elle quantifie leur degré de dépendance en calculant des p-valeurs, de la manière suivante.

Description de la méthode UE initiale.

- La méthode commence par un pré-traitement des données ; les données sont discrétisées (*binning*) et réduites à des évènements binaires (*clipping*) de la manière suivante. Pour une longueur arbitraire de pas δ (généralement de l'ordre de plusieurs millisecondes), le temps de l'expérience est discrétisé en petites plages de temps, appelées "bins", de longueur δ, et à chaque bin est associé un "1" si au moins un spike apparaît pendant le bin, et un "0" sinon. Le train de spikes peut alors être vu comme un vecteur aléatoire à coordonnées dans $\{0, 1\}$, et de dimension finie égale au nombre de bins.

- On parle de coïncidence dans un bin entre les neurones du motif fixé lors d'un essai lorsque tous les trains de spikes correspondants ont un "1" dans ce bin. Ensuite, le nombre total de coïncidences observé n_{emp} correspond à la somme sur les essais du nombre de bins contenant une coïncidence entre les neurones du motif. Une définition plus précise dans le cas de $L = 2$ neurones est donnée dans la définition 2.2.1 dans le chapitre 2.

- Comme nous l'avons déjà mentionné, l'idée générale est de comparer le nombre total de coïncidences observé à celui moyen sous l'indépendance des activités des neurones. Pour cela, le nombre total de coïncidences moyen est approché par une fonction de la probabilité jointe que tous les neurones du motif aient un "1" dans un bin. Sous indépendance, cette loi jointe est égale au produit de ses marginales, et donc, estimer ces marginales mène à un bon estimateur du nombre total de coïncidentes moyen n_{exp}.

- En supposant par ailleurs que ce nombre moyen suit une loi de Poisson, la p-valeur est donc définie comme la probabilité qu'une variable aléatoire de loi de Poisson de paramètre n_{exp} soit supérieure ou égale au nombre total observé n_{emp}.

  Ensuite, le test rejette l'indépendance si la p-valeur est soit inférieure ou égale au niveau prescrit α (signifiant un excès de synchronisations, et donc une corrélation excitatrice), soit supérieur ou égal à $1-\alpha$ (signifiant un manque significatif de coïncidences, et donc une corrélation inhibitrice).

- Finalement, afin de détecter les moments où les neurones du motif fixé coordonnent leur activité, le test est appliqué simultanément sur de nombreuses plages de temps, et les synchronisations détectées correspondent aux fenêtres pour lesquelles le test rejette l'indépendance.

Plusieurs améliorations de la méthode initiale ont été introduites et sont en partie décrites dans l'introduction du chapitre 2.
Premièrement, la discrétisation des données peut impliquer une perte drastique d'information pour la détection des synchronisations (voir [72]), étant donné que le nombre de coïncidences dépend fortement de la longueur et de la position de chaque bin. Notamment, les coïncidences impliquant un spike du premier neurone proche de la fin d'un bin, et un spike du second neurone proche du début du bin suivant ne peuvent pas être détectées, aussi proches que soient les temps de spikes. Afin d'éviter une telle perte d'information, une nouvelle notion de coïncidence, sans discrétisation grossière, appelée *multiple shift coincidence count*, est introduite par Grün et al. dans [72] pour des données discrétisées plus finement, et a été généralisée par la notion de *coïncidence avec délai*, notée φ_{δ}^{coinc}, entre deux processus ponctuels par Tuleau-Malot et al. dans [170] (voir la définition 2.2.2 dans le chapitre 2 pour plus de détails) et ensuite entre trois neurones ou plus par Chevallier et Laloë dans [35].

Deuxièmement, l'hypothèse de loi de Poisson concernant le nombre de coïncidences ne semble pas réaliste.

Troisièmement, dans la théorie des tests multiples, la multiplicité des tests doit être prise en compte. Par exemple, si l'on applique simultanément m tests indépendants de tailles individuelles égales à α, alors, globalement, la probabilité de rejeter au moins une hypothèse nulle est égale à $(1-(1-\alpha)^m)$, ce qui équivaut à $m\alpha$ lorsque α tend vers zéro. En particulier, c'est largement supérieur à α si le nombre de tests m est grand. Ainsi, chaque niveau individuel doit être corrigé. Il existe de nombreuses propositions de corrections dans la littérature. Parmi

elles, nous avons appliqué la célèbre procédure de Benjamini et Hochberg [16], rappelée dans la section 0.3.4.

Tuleau-Malot et al. [170] proposent une méthode justifiée théoriquement pour détecter les synchronisations entre deux processus ponctuels (généralisée dans [35] à trois neurones ou plus) qui ne souffre pas de ces inconvénients. Cette méthode, appelée *Multiple Tests based on a Gaussian Approximation of the Unitary Events* (MTGAUE), est basée sur une approximation gaussienne du nombre de coïncidences avec délai recentré et renormalisé. Plus précisément, supposons que nous observons sont n copies $(X_1, \ldots, X_n)$ i.i.d. d'un couple de processus ponctuels $X = (X^1, X^2)$.

Description de la méthode MTGAUE.

- Fixons une fenêtre $W = [a, b]$ et calculons la moyenne empirique du nombre de coïncidences (avec délai) φ_δ^{coinc} sur W défini comme

$$\bar{Z}_n = \frac{1}{n} \sum_{i=1}^{n} \varphi_\delta^{coinc}(X_i^1, X_i^2).$$

- Si X^1 et X^2 sont deux processus de Poisson homogènes indépendants d'intensités λ^1 et λ^2 (voir la définition en section 0.3.2), alors le nombre moyen de coïncidences entre X^1 et X^2 peut être exprimé en fonction de λ^1 et λ^2, et est noté dans la suite

$$\mathbb{E}\left[\varphi_\delta^{coinc}(X^1, X^2)\right] = m(\lambda^1, \lambda^2).$$

  L'étape de plug-in consiste à remplacer les paramètres inconnus λ^1 et λ^2 par leurs estimateurs sans biais $\hat{\lambda}^1$ et $\hat{\lambda}^2$ définis par

$$\hat{\lambda}^j = \frac{1}{n(b-a)} \sum_{i=1}^{n} N_{X_i^1}([a, b]),$$

  où $N_{X_i^1}([a, b])$ désigne le nombre de points de X_i^1 dans $[a, b]$.

- Ils montrent alors que sous l'hypothèse nulle, leur statistique de test T_n, correspondant à $(\bar{Z}_n - m(\hat{\lambda}^1, \hat{\lambda}^1))$ correctement normalisé, converge en loi vers la loi normale centrée et renormalisée $\mathcal{N}(0, 1)$. Ainsi, le test bilatéral rejette l'indépendance sur la fenêtre W si $|T_n|$ est strictement supérieur au quantile d'ordre $(1 - \alpha/2)$ de $\mathcal{N}(0, 1)$.

- Finalement, ainsi que pour la méthode UE initiale, afin de détecter lorsque les neurones coordonnent leur activité, le test est appliqué simultanément sur plusieurs fenêtres. Néamoins, afin de prendre en compte la multiplicité des tests, ils appliquent la procédure de Benjamini et Hochberg (voir la section 0.3.4).

Cependant, cette méthode, ainsi que la méthode UE initiale, est toujours basée sur des hypothèses très restrictives sur les trains de spikes, telles que des hypothèses fortes de stationnarité, et risquent donc de ne pas fonctionner correctement sur des données expérimentales.

Pour éviter ces hypothèses de stationnarité, plusieurs procédures ont également été introduites dans la littérature, basées sur des approches par rééchantillonnage que nous classons ci-dessous en deux grandes familles.

La première se base sur des hypothèses de modèles non-paramétriques, tels que les processus de Poisson inhomogènes par exemple. L'intensité de tels processus est estimée de manière

non-paramétrique, puis des processus d'intensité estimée sont simulés afin de reconstruire la loi sous $(\mathcal{H}_0)$. Cette approche est un cas particulier du bootstrap paramétrique. Elle est utilisée, par exemple, dans [175, 176, 103], se basant sur des modèles plus complexes ayant des intensités conditionnelles dépendant du passé du processus, tels que les processus inhomogènes à intervalles markoviens, les processus gamma, ou des modèles log-linéaires.

La seconde approche comprend toutes les méthodes basées sur du rééchantillonnage permettant de reconstruire la loi sous $(\mathcal{H}_0)$ sans aucune hypothèse de modèle. Ces méthodes comprennent toute la famille des méthodes de substitution des données (appelées *surrogate data methods* en anglais), répertoriées dans [120], telles que la randomisation des temps de spikes ou les méthodes de *dithering* (pouvant être traduit par *flottement* en français) telles que le décalage des temps de spikes, ou des trains de spikes (voir [121]) qui créent de nouvelles données à partir des anciennes, et détruisent les structures de dépendance éventuelles en modifiant la structure interne des trains de spikes à travers le temps. Cependant, nous ne connaissons pas de structures mathématiques qui permettraient de légitimer théoriquement de telles méthodes, à part peut-être les processus de renouvellement.

Nous pouvons également considérer d'autres méthodes de substitution des données, basées sur des approches de type bootstrap introduites rapidement par Ventura dans [175] et basées sur l'*histogramme joint des temps peri-stimulus* (jPSTH) (voir [3]). Plus précisément, considérons la discrétisation des données introduit dans la première étape de la méthode UE initiale, mais avec une longueur de bin $\varepsilon > 0$ suffisamment petite de sorte qu'aucun bin ne puisse contenir plus d'un spike. Alors, l'*histogramme des temps peri-stimulus* (PSTH) d'un neurone est la somme bin par bin sur les essais des trains de spikes discrétisés divisé par le nombre d'essais. Il constitue un estimateur naturel du taux de décharge instantané du neurone, à savoir le nombre moyen (théorique) de spikes par unité de temps. Le jPSTH correspondant à deux neurones est le PSTH du produit bin par bin des trains de spikes discrétisés individuels. Afin de résoudre l'effet de non-stationnarité, la statistique de test basée sur le jPSTH et les produits des PSTH marginaux, dépendent du temps t. Remarquons que cette statistique de test peut s'écrire comme un cas particulier de la notre, à savoir $U_{n,h_{\varphi_t}}$ définie par (0.1.3) et (0.1.4), où $\varphi_t(X^1, X^2) = \mathbb{1}_{N^1_X([t,t+\varepsilon[)=1,\ N^2_X([t,t+\varepsilon[)=1}$ pour chaque t correspondant à un bin. Sa distribution sous l'hypothèse d'indépendance est obtenue grâce à une méthode de bootstrap non-paramétrique (avec ou sans approximation par Monte Carlo), consistant à rééchantillonner indépendamment pour chaque neurone, comme introduit dans la section 0.1.3. Afin de résoudre le problème de multiplicité dû à la dépendance en temps, le test rejette l'indépendance lorsque la statistique de test, en tant que courbe, sort de l'enveloppe estimée sous l'hypothèse nulle. Cependant, au vue du chapitre 1, et afin de justifier entièrement cette approche d'un point de vue théorique, il faudrait démontrer que cette approche par bootstrap ne permet pas seulement de reconstruire la loi sous $(\mathcal{H}_0)$ des $U_{n,h_{\varphi_t}}$ pour chaque bin t, mais de $\sup_t \left|U_{n,h_{\varphi_t}}\right|$. Par ailleurs, le choix de la longueur des bins ε est très délicat. En effet, si ε est trop petit, $\varphi_t(X^1, X^2)$ risque d'être nul pour la plupart des bins, alors que si ε est trop grand, la méthode risque de souffrir d'une perte d'information comme pour n'importe quelle discrétisation grossière de type *binning*.

Une autre méthode largement répandue est la méthode *trial-shuffling* développée par Pipa et al. [137, 138], détruisant éventuelles structures de dépendance en cassant les couples de trains de spikes à travers les essais. Intuitivement, cette approche consiste à tirer avec remise dans tous les couples de trains de spikes évitant ceux d'un même essais, c'est-à-dire avec les notations précédentes $\{(X^1_i, X^2_j)\}_{i\neq j}$ (voir la section 2.3.2 dans le chapitre 2 pour plus de détails). Même si cette approche est généralement considérée comme une méthode par permutation dans la littérature en neurosciences, elle ne correspond pas à celle introduite

dans cette thèse (voir la section 0.1.3). En effet, d'une part, l'échantillon permuté peut contenir un couple de trains de spikes provenant d'un même essai (dès lors que la permutation aléatoire a un point fixe), ce qui est interdit pour le *trial-shuffling* par définition, et d'autre part, ce dernier consiste à tirer *avec* remise, ce qui fait qu'il se rapproche plutôt des méthodes par bootstrap que de celles par permutation. Plus précisément, cette méthode est basée sur une approche de type bootstrap non-paramétrique (comme celle dans [175], et contrairement à celles dans [103, 176]), et ne nécessite donc pas d'hypothèse de modèle. Cependant, elle est basée sur une discrétisation des données et risque donc également de souffrir d'une perte d'information, comme la méthode UE initiale. De plus, au vue de l'analyse du problème de recentrage faite dans le chapitre 2, les méthodes par bootstrap appliquées à des statistiques non-centrées ne permettent généralement pas d'approcher la loi sous $(\mathcal{H}_0)$. En particulier, le *trial-shuffling* considère directement le nombre de coïncidences (non-recentré), et ne permet donc pas de reconstruire la loi sous $(\mathcal{H}_0)$ souhaitée (voir le chapitre 2).

0.3.2 Modélisation par des processus ponctuels

Les processus ponctuels apparaissent comme des outils de modélisation naturels de l'activité neuronale (voir, par exemple, l'introduction aux processus ponctuels dans le livre de Kass et al. [102, Chapitre 19]). En effet, même si les trains de spikes enregistrés expérimentalement sont discrétisés en temps à cause de la résolution d'enregistrement, et appartiennent donc à des espaces de dimension finie, cette dernière étant tellement grande (entre dix mille et un million) qu'il n'est ni réaliste, ni raisonnable de les modéliser par des vecteurs de dimension finie, et les processus ponctuels semblent être mieux adaptés.

Une brève introduction aux processus ponctuels

Les processus ponctuels ont été largement étudiés au cours de ces dernières décennies dans de nombreux domaines tels que l'économie, l'épidémiologie, la sismologie, les télécommunications, les neurosciences, etc. Une introduction complète à la théorie des processus ponctuels est faite dans le livre de Daley et Vere-Jones [38].
Un processus ponctuel X sur un ensemble mesurable I est une partie dénombrable aléatoire de I. Notons $\mathcal{X}$ l'ensemble de toutes ses valeurs possibles. Pour des raisons biologiques, nous ne considérons que les processus ponctuels presque sûrement localement finis. Ils sont caractérisés par le fait que, quel que soit le sous-ensemble J de I, le nombre de points de X dans J, noté $N_X(J)$ est presque sûrement fini. En particulier, de tels processus ne peuvent pas avoir de point d'accumulation.
Pour tout x dans $\mathcal{X}$, définissons sa mesure de comptage associée comme $dN_x = \sum_{u\in x} \delta_u$. En particulier, elle vérifie pour toute fonction mesurable à valeurs réelles f sur I,

$$\int_I f(s) dN_x(s) = \sum_{u\in x} f(u).$$

Dans la suite, puisque nous nous intéressons aux processus ponctuels indexés par le temps, nous nous concentrons sur le cas particulier où I est de la forme $[0,T]$ où $0 < T \leq +\infty$. Dans ce cas, un processus ponctuel X peut être identifié à son processus de comptage, défini par

$$N_X : \begin{pmatrix} I & \longrightarrow & \mathbb{R} \\ t & \longmapsto & \int_0^T \mathbb{1}_{s\leq t}\, dN_X(s) \end{pmatrix}.$$

En particulier, $N_X(t)$ est une variable aléatoire comptant le nombre de points de X dans $[0,t]$. Le processus $(N_X(t))_{t\geq 0}$ est positif, croissant, constant par morceaux, chaque saut étant égal à

un et apparaissant à chaque point du processus ponctuel associé X. Il est par ailleurs continu à droite et limité à gauche (càdlàg). À travers cette identification l'ensemble des processus ponctuels $\mathcal{X}$ peut être muni d'une distance $d_{\mathcal{X}}$ issue de la topologie de Skorokhod grâce à son plongement dans l'ensemble des fonctions càdlàg (voir [22] ou Appendix A.2.2).

La loi d'un processus ponctuel peut être définie de différentes manières. Premièrement, elle peut être représentée par la mesure de probabilité définie sur les ensembles boréliens de $(\mathcal{X}, d_{\mathcal{X}})$, comme c'est le cas dans cette thèse (voir la section 0.1.2). Cependant, cette définition ne permet pas de caractériser simplement les processus ponctuels usuels.
Un autre outil plus classique pour caractériser la loi d'un processus ponctuel, est l'intensité conditionnelle $\lambda(t)$ sachant le passé. Essentiellement, elle dépend de l'histoire du processus jusqu'au temps t, codée dans la filtration naturelle du processus définie comme la tribu $\mathcal{F}_{t^-}$ engendrée par $\{N_X(s), s < t\}$. Intuitivement, λ est une fonction mesurable positive telle que $\lambda(t)dt$ est la probabilité conditionnelle de trouver un nouveau point dans $[t, t + dt)$ sachant $\mathcal{F}_{t^-}$, c'est-à-dire

$$\lambda(t)dt \approx \mathbb{E}\left[dN_X(t)|\mathcal{F}_{t^-}\right].$$

Cette thèse n'est pas destinée à donner les définitions précises de tels outils, et nous recommandons les livres de Brémaud [28] ou de Daley et Vere-Jones [38] pour plus de détails. Un des grands avantages de cette définition est que, étant donnée l'intensité conditionnelle, il est possible de construire un processus ponctuel de telle intensité de manière simple grâce à la procédure de *thinning* de Ogata [129] décrite dans l'annexe A.2.1. En neurosciences, cette notion correspond au taux de décharge instantané, qui décrit également la probabilité infinitésimale qu'un neurone décharge.

Quelques exemples typiques

Processus de Poisson homogènes. L'exemple le plus simple de processus ponctuels est celui des *processus de Poisson homogènes* (voir par exemple [157, Définition 5.1]).

Definition 0.3.1. Un processus de Poisson homogène X de paramètre $\lambda > 0$ sur I est défini par

- $N_X(0) = 0$,
- X est à accroissements indépendants, ce qui signifie que, pour tout entier naturel l, et tous ensembles mesurables deux à deux disjoints $A_1, \ldots, A_l$ de I, $N_X(A_1), \ldots, N_X(A_l)$ sont des variables aléatoires indépendantes,
- pour tout $s > 0$, le nombre de points dans n'importe quel intervalle de longueur s suit une loi de Poisson de paramètre λs, c'est-à-dire que pour tout $t > 0$ tel que $[t, t + s)$ est inclus dans I, pour tout entier naturel k,

$$\mathbb{P}(N_X([t, t + s)) = k) = \frac{(\lambda s)^k}{k!} \exp(-\lambda s).$$

Les processus de Poisson présentent de nombreuses propriétés avantageuses. En particulier, ils sont à accroissements stationnaires, ce qui signifie que la loi du nombre de points appartenant à un intervalle quelconque ne dépend que de la longueur de ce dernier, et non de son positionnement dans I. De plus, l'intensité conditionnelle d'un processus de Poisson est déterministe, constante et égale à son paramètre λ. Cette dernière propriété simplifie notamment l'étude de tels processus et c'est pour cette raison qu'ils sont très souvent utilisés dans la littérature en neurosciences (voir par exemple [67, 170, 35]) même s'ils sont trop simples pour être réalistes.

Processus de Poisson inhomogènes. Une première généralisation des processus de Poisson homogènes consiste à supprimer la propriété de stationnarité, ce qui nous amène aux *processus de Poisson inhomogènes*.

Definition 0.3.2. Un processus de Poisson inhomogène X de mesure moyenne μ ne contenant aucun atome sur I est défini par

- X est à accroissements indépendants (voir la définition 0.3.1),
- pour tout ensemble mesurable A de I, le nombre de points dans A suit une loi de Poisson de paramètre $\mu(A)$, c'est-à-dire que pour tout entier naturel k,

$$\mathbb{P}(N_X(A) = k) = \frac{\mu(A)^k}{k!} \exp\left(-\mu(A)\right).$$

Généralement, la mesure μ est supposée être absolument continue par rapport à la mesure de Lebesgue, et s'écrit donc $\mu(dt) = \lambda(t)dt$. La fonction λ est alors appelée intensité du processus.

D'une part, remarquons que l'intensité d'un processus de Poisson inhomogène est égal à son intensité conditionnelle. D'autre part, si cette intensité est constante, nous retrouvons un processus de Poisson homogène.

L'avantage principal des processus de Poisson (homogènes ou non) est que leur intensité conditionnelle est déterministe, et ne dépend pas du passé. Cette propriété les rend attractifs d'un point de vue mathématique. Dans la littérature, ils sont souvent utilisés pour modéliser, par exemple, les temps d'émission des particules radioactives, les appels téléphoniques ou les temps d'arrivée de clients dans une banque. Cependant, ils ne sont pas réalistes en neurosciences, étant donné qu'il ne permettent pas de modéliser des caractéristiques biologiques telles que les périodes réfractaires, qui sont de brefs délai après chaque spike pendant lesquels le neurone ne peut pas décharger.

Processus de Hawkes et modèles à interactions poissonniennes. Les *processus de Hawkes* constituent une nouvelle généralisation permettant de modéliser des particularités telles que l'auto-excitation, ou au contraire, l'auto-inhibition (comme la période réfractaire en neurosciences). Introduits par Hawkes [78] pour modéliser les répliques des tremblements de terre en sismologie (voir [130, 177]), les processus de Hawkes ont été appliqués dans des domaines variés tels que la génomique (voir [73, 149]), la finance (voir [11, 12]) et plus récemment les neurosciences (voir [113, 134, 147, 148, 160]).

Definition 0.3.3. Un processus de Hawkes univarié est un processus ponctuel X d'intensité conditionnelle de la forme

$$\lambda(t) = \Phi\left(\int_0^t h(t-s)dN_X(s)\right) = \Phi\left(\sum_{u\in X, u<t} h(t-u)\right), \tag{0.3.1}$$

où Φ est une fonction positive, et $h : \mathbb{R}_+ \to \mathbb{R}$ est appelée fonction d'auto-interaction. Tout au long de cette thèse, nous considérerons Φ égal à $\Phi(\cdot) = \max\{0, \mu + \cdot\}$, où $\mu > 0$ désigne la contribution spontanée.

La dépendance en l'histoire du processus est codée dans le terme d'intégrale/somme dans (0.3.1). En particulier, chaque point du processus u apparaissant avant le temps t donne une "opinion" sur l'apparition d'un nouveau point en t à travers la fonction h en fonction de son âge $t-u$. En particulier, si $h(t-u)$ est strictement positif, alors l'"opinion" de u augmente la probabilité instantanée de trouver un nouveau point en t, et a donc un effet excitateur, alors que si c'est strictement négatif, alors u a un effet inhibiteur. En particulier, si h est très fortement négative sur un petit voisinage de zéro, disons $[0,r]$, de sorte que pour tout point u du processus, $\mu + \int_0^t h(t-s)dN_X(s)$ est négatif pour tout t appartenant à $[u,u+r]$, alors l'intensité est égale à zéro sur chaque intervalle $[u,u+r]$ et donc, aucun point ne peut apparaître à une distance inférieure à r d'un point existant. Ceci modélise exactement la période réfractaire. Finalement, tous les points u du processus tels que $h(t-u)=0$ n'ont "aucune opinion" sur l'apparition d'un nouveau point au temps t. En particulier, si h est identiquement nulle, alors, nous retrouvons un processus de Poisson homogène.

Les interactions entre neurones étant la motivation principale de l'analyse des synchronisations, les processus de Hawkes multivariés (dépendants) ont également été introduits dans la littérature.

Definition 0.3.4. Un processus de Hawkes bivarié est un couple (X^1, X^2) de processus de Hawkes (interagissant) où pour $i=1,2$, l'intensité conditionnelle λ^i de X^i est de la forme

$$\lambda^i(t) = \Phi_i\left(\sum_{j=1}^{2}\int_0^t h_{j\to i}(t-s)dN_{X^j}(s)\right),$$

où Φ_i est une fonction positive (prise comme dans la définition 0.3.3), et $h_{j\to i}:\mathbb{R}_+\to\mathbb{R}$ est appelée fonction d'auto-interaction si $j=i$ et fonction d'interaction sinon.

En plus de la dépendance en sa propre histoire, chaque composante d'un processus de Hawkes bivarié (X^1,X^2), disons X^1, dépend de celle de l'autre processus, disons X^2, à travers la fonction d'interaction $h_{2\to1}$. En particulier, en neurosciences, si la fonction $h_{2\to1}$ est positive, alors le neurone 2 a tendance à exciter l'activité du neurone 1, alors que si elle est négative, alors le neurone 2 a tendance à inhiber l'activité du neurone 1. Cette définition peut être généralisée à plus de deux processus, et donne lieu aux processus de Hawkes dits multivariés.

Le *modèle à interactions poissonniennes* introduit par Sansonnet et Tuleau-Malot dans [160] est un autre modèle pour les trains de spikes rencontré dans l'analyse des synchronisations en neurosciences. C'est un cas particulier des processus de Hawkes bivariés (X^1,X^2), où les deux fonctions d'auto-interaction $h_{1\to1}$ et $h_{2\to2}$, et la fonction d'interaction $h_{2\to1}$ sont identiquement nulles. Dans ce cas le processus "parent" X^1 est simplement un processus de Poisson homogène, et le processus "enfant" X^2 dépend uniquement de l'histoire du processus parent. Tester l'indépendance revient donc à tester la nullité de la fonction $h_{1\to2}$.

Par ailleurs, un processus de Hawkes pour lequel la fonction Φ dans la définition 0.3.3 est la fonction exponentielle, est un cas particulier des *modèles linéaires généralisés* (voir par exemple [136, 169]).

Cependant, comme nous l'avons déjà mentionné, aucun de ces modèles, ni aucun autre, n'est communément accepté aujourd'hui dans la littérature en neurosciences. Néanmoins, nous utilisons ces modèles classiques dans nos études par simulation.

0.3.3 Une brève introduction aux tests multiples

Afin de répondre à la problématique de détection de synchronisations, l'approche classique consiste à tester l'indépendance simultanément sur plusieurs fenêtres glissantes. Cela nécessite la théorie des tests multiples que nous introduisons rapidement dans cette section. La problématique de tests multiples apparaît dès lors que l'on souhaite tester simultanément m hypothèses nulles, disons $(\mathcal{H}_{0,1}), \dots, (\mathcal{H}_{0,m})$ tout en contrôlant l'erreur globale par un certain niveau α.
Une procédure de tests multiples est une fonction des observations renvoyant un sous-ensemble de $\{1, \dots, m\}$ correspondant aux indices des hypothèses nulles rejetées. Dans la suite, nous appellerons un *positif* une hypothèse nulle déclarée significative, et donc rejetée, et au contraire un *négatif* une hypothèse nulle non significative au vue des données, et donc acceptée. Un *faux positif* (respectivement *faux négatif*) désigne donc une hypothèse nulle rejetée (respectivement acceptée) à tort. Ainsi, par analogie avec la théorie des tests d'hypothèses, le risque de première espèce d'une procédure de tests multiples est étroitement lié au nombre de faux positifs.
Considérons les notations très répandues, popularisées par Benjamini et Hochberg dans [16], et rappelées dans la table 1.

	Hypothèses nulles rejetées	Hypothèses nulles acceptées	Total
Hypothèses vraies	U	V	m_0
Hypothèses fausses	T	S	$m - m_0$
Total	$m - R$	R	m

Table 1 – Nombre d'erreurs commises lors du test multiple de m hypothèses nulles.

En particulier, remarquons que le nombre de positifs R est une variable aléatoire observable alors que le nombre de faux positifs V et le nombre d'hypothèses nulles vraies m_0 sont tous deux inconnus.

Contrairement aux tests d'hypothèses (simples), il existe plusieurs notions de risque de première espèce d'une procédure de tests multiples. Historiquement, la première définition a avoir été introduite est le *Family Wise Error Rate* (FWER) donné par

$$\text{FWER} = \mathbb{P}(V \geq 1)\,,$$

c'est-à-dire la probabilité que la procédure détecte au moins un faux positif. En particulier, comme nous l'avons mentionné dans la section 0.3.1, si la multiplicité des tests n'est pas prise en compte, et que chaque test est appliqué au niveau individuel α, alors, le FWER est contrôlé par $\alpha \times m$. En effet, par la propriété de σ-additivité des mesures de probabilité,

$$\text{FWER} \leq \sum_{(\mathcal{H}_{0,i}) \text{ vraie}} \mathbb{P}(\text{rejeter à tort } (\mathcal{H}_{0,i})) \leq m_0\alpha \leq m\alpha.$$

Ainsi, une correction naturelle, introduite par Bonferroni (voir [87]) consiste à diviser chaque niveau individuel par le nombre de test, ce qui donne α/m. Cependant, lorsque le nombre de tests est très grand, les niveaux corrigés sont tellement petits que la procédure risque de ne plus rejeter aucune hypothèse nulle et devient donc très conservative.
Pour éviter cela, une autre notion de risque de première espèce a été popularisée par Benjamini et Hochberg [16]. Le *taux de faux positifs* (FDR pour *False Discovery Rate* en anglais) est

défini par

$$\mathrm{FDR} = \mathbb{E}\left[\frac{V}{R}\mathbb{1}_{R>0}\right].$$

Il représente le taux moyen de faux positifs parmi les positifs. Le FDR est plus faible que le FWER au sens de l'inégalité FDR $\leq$ FWER (avec égalité lorsque $m_0 = m$, ce qui signifie que toutes les hypothèses nulles sont vraies). En particulier, le FDR mène à des procédures généralement moins conservatives puisqu'il est moins strict sur le nombre de faux positifs autorisés, et constitue l'une des premières alternatives au FWER communément acceptée dans de nombreux domaines (tels que la génomique, les études cliniques, les réseaux d'interactions, etc).

De nombreuses procédures de tests multiples ont été développées afin de contrôler le FWER, le FDR, et même d'autre notions de risque de première espèce (voir, par exemple, le livre Dudoit et van der Laan [47]). Parmi elles, nous nous sommes concentrés sur la célèbre procédure de Benjamini et Hochberg [16] décrite ci-dessous, basée sur les p-valeurs $\{p_1, \ldots, p_m\}$ correspondant aux hypothèses nulles $\{(\mathcal{H}_{0,1}), \ldots, (\mathcal{H}_{0,m})\}$:

- D'abord, ordonnons les p-valeurs $p_{(1)} \leq \cdots \leq p_{(m)}$, et notons $(\mathcal{H}_{0,(i)})$ l'hypothèse nulle correspondant à $p_{(i)}$.
- Ensuite, considérons $\hat{k} = \max\left\{i \ ; \ p_{(i)} \leq \frac{i\alpha}{m}\right\}$.
- Finalement, rejetons les hypothèses nulles correspondant aux $\hat{k}$ plus petites p-valeurs, à savoir $(\mathcal{H}_{0,(1)}), \ldots, (\mathcal{H}_{0,(\hat{k})})$.

La popularité de cette méthode vient de sa simplicité d'application, son efficacité à traiter tous types de problématiques dans de nombreux domaines très variés, et sa capacité à contrôler le FDR sous certaines hypothèses. Par exemple, il est indiqué dans [16] que :

> Si les p-valeurs sont indépendantes, alors la procédure de Benjamini et Hochberg contrôle le FDR au niveau α.

L'hypothèse d'indépendance entre les p-valeurs étant plutôt restrictive, Benjamini et Yekutiely [17] l'ont relâché à une propriété de dépendance positive. Plus précisément, la propriété de dépendance positive sur un sous-ensemble I_0 de $\{1, \ldots, m\}$ (PRDS pour *Positive Regression Dependency on a Subset* en anglais) est vérifiée par un vecteur aléatoire $(\xi_1, \ldots, \xi_m)$ si pour tout ensemble croissant D dans $[0,1]^m$ (caractérisé par "si $x \in D$ et $x \leq y$ alors $y \in D$"), et si pour tout i dans I_0, l'application $u \mapsto \mathbb{P}((p_1, \ldots, p_m) \in D | p_i = u)$ est croissante. En particulier, ils démontrent que :

> Si les p-valeurs vérifient la propriété PRDS sur le sous-ensemble I_0 contenant les indices des hypothèses nulles vraies, alors la procédure de Benjamini et Hochberg contrôle le FDR au niveau α.

Remarquons que Blanchard et Roquain fournissent une preuve beaucoup plus simple et élégante de ce résultat en tant que cas particulier d'un résultat plus général dans [23].

0.3.4 La méthode Permutation UE et application à de vraies données

Rappelons le raisonnement général (comme dans [170]) pour répondre à la problématique de détection de synchronisations entre deux neurones : la première étape consiste à construire un test d'indépendance entre deux neurones s'appliquant sur une petite fenêtre de temps, et la seconde étape consiste à appliquer une procédure de tests multiples détectant les plages sur lesquelles les neurones synchronisent leur activité.

Contribution de cette thèse

Première étape : tests individuels. Considérons une fenêtre de temps $W = [a, b]$ incluse dans le temps de l'expérience $[0, T]$. Avant d'appliquer les tests d'indépendance par bootstrap ou par permutation que nous avons introduit dans la section 0.1.3, nous devons choisir la fonction φ dans la définition de la statistique de test dans le *cas linéaire* $h = h_\varphi$ (voir (0.1.3) et (0.1.4)). Un choix bien adapté à la problématique de détection de synchronisations repose sur la notion de coïncidences avec délai introduite par Tuleau-Malot et al. dans [170]. Plus précisément, le nombre de coïncidences avec délai δ, entre deux processus ponctuels X^1 et X^2 (modélisant deux trains de spikes enregistrés simultanément) sur l'intervalle $[a, b]$ est défini par

$$\varphi_\delta^{coinc}(X^1, X^2) = \int_a^b \int_a^b \mathbf{1}_{|u-v|\leq\delta} dN_{X^1}(u) dN_{X^2}(v).$$

Il compte le nombre de fois qu'un spike du premier neurone et un spike du second neurone apparaissent à une distance inférieure ou égale à δ. Remarquons que φ_δ^{coinc} ne peut pas s'écrire comme un produit, et donc les tests par bootstrap ou par permutation rencontrés dans la littérature en statistique, tels que ceux de Romano [155] et de van der Vaart et Wellner [173], ne sont pas applicables ici.

Intuitivement, les tests unilatéraux à droite par bootstrap ou par permutation détectent la fenêtre W si le nombre de coïncidences sur W est significativement trop grand comparé à ce qui est attendu sous indépendance.

Une étude par simulation (présentée dans la section 1.5 du chapitre 1) a été effectuée afin de vérifier la validité d'un point de vue pratique des tests d'indépendance par bootstrap et par permutation introduits dans la section 0.1.3 basés sur le nombre de coïncidences avec délai. De plus, leurs performances, en termes de risques de première et de seconde espèces, sont comparées à celles de méthodes classiques en neurosciences, telles que le *trial-shuffling* (voir [137, 138]), ou les tests individuels de la méthode MTGAUE (voir [170]), sur différents types de processus ponctuels. En quelques mots, comme nous pouvions nous y attendre, le test par permutation est le seul à contrôler non-asymptotiquement son risque de première espèce, et il semble être aussi performant que les autres en termes de risque de seconde espèce. Pour cela, c'est le seul que nous considérons dans l'étape de tests multiples. Par ailleurs, nous pouvons voir que nos tests par bootstrap et par permutation semblent être toujours aussi performants que le meilleur des tests étudiés, sans souffrir de restrictions au processus de Poisson homogènes (comme la méthode MTGAUE) ou du manque de recentrage (comme la méthode *trial-shuffling*).

Seconde étape: tests multiples. Nous construisons dans le chapitre 2 une procédure de test multiple basée sur celle de Benjamini et Hochberg permettant de détecter à la fois les corrélations excitatrices et celles inhibitrices. L'idée est d'appliquer simultanément les tests unilatéraux à droite et à gauche sur chaque fenêtre.
Ainsi, étant donnée une famille de K fenêtres non-nécessairement disjointes, recouvrant tout l'intervalle $[0, T]$, nous calculons sur chaque fenêtre W les p-valeurs p_W^+ et p_W^- correspondant aux tests unilatéraux à droite et à gauche par permutation basés sur le nombre de coïncidences avec délai, et appliquons la procédure de Benjamini et Hochberg à la famille des $m = 2K$ p-valeurs.
À ce jour, nous n'avons quasiment aucune garantie théorique du contrôle du FDR par notre procédure. Cependant, elle semble donner de bons résultats d'un point de vue pratique, comme nous le pouvons le voir dans l'étude par simulation faite dans la section 2.4.2 du chapitre 2.

Application à de vraies données. Afin de compléter cette étude, nous appliquons cette procédure à des données expérimentales ayant déjà été étudiées dans la littérature par Riehle, Grammont et bien d'autres (voir [150, 64, 151, 170]).

Le protocole expérimental. Des micro-électrodes sont implantées dans le cortex moteur d'un singe Rhésus entrainé à pointer du doigt dans plusieurs directions. L'animal est assis face à sept diodes lumineuses (LED) tactiles ; une, placée au centre, est entourée de six autres équidistantes. L'expérience débute lorsque le singe touche la LED centrale. Après un délai de 500ms, une des LEDs périphériques s'allume en vert, indiquant la cible du mouvement à venir ; c'est le *signal préparatoire* (PS). Ensuite, après un délai soit de 600ms (avec probabilité 0.3) ou de 1200ms sinon, la LED périphérique allumée devient rouge indiquant au singe de la pointer du doigt ; c'est le *signal de réponse* (RS). Le singe est récompensé par une goute de jus après chaque bonne réponse, et une pause entre chaque essai est respectée, permettant de justifier de l'hypothèse que ces essais sont indépendants et identiquement distribués. Une description plus complète de l'expérience est faite dans la section 2.4.2 du chapitre 2, avec des précisions sur la technique d'enregistrement.

Les résultats obtenus (voir la figure 2.9 du chapitre 2) mettent en évidence la capacité à détecter des moments clés de l'expérience de cette nouvelle procédure de test. De plus, au vue de l'étude par simulation, elle semble plus fiable que les autres méthodes (et en particulier, MTGAUE), puisqu'elle ne nécessite pas d'hypothèse de stationnarité. Par conséquent, elle détecte moins de faux positifs, dont, notamment, les fenêtres sur lesquelles les données sont fortement non-stationnaires.

Organisation de cette thèse

En résumé, voici le contenu de ce manuscrit de thèse.

Le chapitre 1 contient la construction des tests d'indépendance par bootstrap et par permutation. Il est consacré à l'étude asymptotique de ces tests dans un cadre de processus ponctuels. Plus précisément, nous montrons que les lois conditionnelles des statistiques de test bootstrappée et permutée sachant l'observation convergent en distance de Wasserstein d'ordre 2 vers la loi asymptotique de la statistique de test sous l'hypothèse nulle ($\mathcal{H}_0$), et ce que l'observation vérifie ou non ($\mathcal{H}_0$). En particulier, cela signifie que les approches par bootstrap et par permutation permettent de reconstruire la loi sous ($\mathcal{H}_0$). Ensuite, nous en déduisons que ces tests ont de bonnes propriétés asymptotiques. Par ailleurs, nous montrons que les hypothèses dans nos résultats sont vérifiées par la plupart des modèles classiques en neurosciences. Ce chapitre contient également une étude par simulation permettant de vérifier la faisabilité de ces méthodes en pratique, et de les comparer à des méthodes existant dans la littérature en neurosciences.

Le chapitre 2 concerne l'application biologique et s'adresse plutôt à un public de neuroscientifiques. Un des objectifs principaux de ce chapitre est de souligner l'importance du recentrage lors de l'application du bootstrap. Plus précisément, nous illustrons par simulation que si la statistique de test n'est pas centrée sous l'hypothèse nulle, alors la loi approchée grâce au bootstrap n'est pas celle souhaitée, à savoir la loi de la statistique de test sous ($\mathcal{H}_0$). Le second objectif de ce chapitre est d'appliquer nos nouvelles procédures de test en neurosciences. Nous définissons les p-valeurs des tests par bootstrap et par permutation basés sur le nombre de coïncidences avec délai (construits dans le chapitre 1)

ainsi que celles de la méthode *trial-shuffling* avec ou sans recentrage. Toutes ces procédures, basées sur ces p-valeurs, sont comparées sur des données simulées. Ensuite, une procédure de tests multiples de type Benjamini-Hochberg basée sur les p-valeurs du test par permutation, appelée *méthode Permutation UE*, est implémentée pour détecter les synchronisations entre deux trains de spikes. La validité d'un point de vue pratique de cette procédure est vérifiée sur des données simulées avant d'être appliquées sur des données expérimentales.

Le chapitre 3 est dédié à l'étude des inégalités de concentration pour les sommes permutées aléatoirement. Grâce à l'inégalité fondamentale de Talagrand pour les permutations aléatoires, et l'astuce de Ledoux pour le passage de la médiane à la moyenne, nous obtenons un inégalité de concentration de type Bernstein. Nous expliquons également la raison pour laquelle nous avons besoin de tels outils sophistiqués pour l'étude des vitesses de séparation uniforme de nos tests par permutation.

Le chapitre 4 est consacré à l'étude des propriétés non-asymptotiques du test par permutation en termes de vitesses de séparation uniforme. Dans la lignée des tests par permutation construits dans le chapitre 1, nous introduisons une procédure de tests agrégés en nous basant sur l'approche par permutation et sur une méthode de seuillage par ondelettes dans un cadre de variables réelles à densité. En particulier, nous étudions sa vitesse de séparation uniforme sur des classes de fonctions régulières, à savoir des espaces de Besov faibles, par rapport à la distance $\mathbb{L}_2$, et obtenons des vitesses qui semblent être optimales et adaptatives au sens du minimax.

Les points n'ayant pas encore étudiés sont détaillés dans un chapitre de conclusion, ouvrant de nouvelles perspectives à la fois dans le domaine de la statistique théorique que dans celui des neurosciences.
Finalement, nous présentons rapidement en annexe les outils mathématiques utilisés tout au long de ce manuscrit, rappelant des résultats fondamentaux appliqués dans cette thèse.
Un résumé long de cette thèse est détaillé dans la section Développement tout à la fin de ce manuscrit.

Les chapitres 1, 2, et 4 étant respectivement publiés, en révision et quasiment soumis à l'heure de l'écriture de ce manuscrit, nous les avons laissés sous leur forme d'article. En particulier, certaines considérations et définitions peuvent être redondantes avec cette introduction.

Introduction (english)

Contents

0.4 Non-parametric tests of independence 47
0.4.1 Two non-parametric approaches: bootstrap and permutation 47
0.4.2 The mathematical testing setting 50
0.4.3 Construction of two non-parametric tests of independence 53
0.5 How to study the performance? 57
0.5.1 Asymptotic study 57
0.5.2 Non-asymptotic study 61
0.5.3 Concentration inequalities: an efficient tool 64
0.6 Application to synchrony detection in neuroscience 67
0.6.1 The neuroscience context and motivation 67
0.6.2 Modeling by point processes 71
0.6.3 A quick introduction to Multiple Testing 75
0.6.4 The Permutation UE method and application on real data 77

0.4 Non-parametric tests of independence

This thesis takes root in the vast domain of non-parametric independence testing. Testing independence is one of the central goals in data analysis and has therefore been vastly investigated in the statistical literature. The present thesis is motivated by considerations in neuroscience that are fully described in Section 0.6 and for which no model is commonly admitted. Therefore, we focus here on the family of tests that are free from the distribution of the observation, called *non-parametric* or *distribution free* tests.

0.4.1 Two non-parametric approaches: bootstrap and permutation

Many methods have been developed in order to construct non-parametric tests of independence. Among them, a first "naive" approach is to consider test statistics whose asymptotic distribution is free from any unknown parameter under the null hypothesis. As an example, one could consider Pearson's [131, 132] historical chi-square test of independence whose statistic has, as its name suggests, an asymptotic chi-square distribution. Hence, given a level of significance α in (0,1), one can easily construct a purely asymptotic test by rejecting the null hypothesis when the test statistic is greater than the $(1-\alpha)$-quantile of the asymptotic distribution, and obtain a test that is asymptotically satisfactory. Yet, in many cases, and especially when applying such tests to biological data (as in Section 0.6.1), the number of

observations is often small, due to economical or biological reasons, and thus such purely asymptotic procedures may not be suitable.

Another huge family of non-parametric tests of independence is the one of *rank tests*, which are based on the ranks of the data. For instance, one can consider the methods of Pitman [140] based on the Pearsonian correlation coefficient, of Hotelling and Pabst [89] or Kendall et al. [105] based on Spearman's rank correlation coefficient, of Kendall [104] based on Kendall's tau coefficient, or of Wolfowitz [179] based on an adapted likelihood ratio method. All these methods rely on relations of order, and are constructed for real-valued random variables or vectors. However, because of the biological motivation detailed in Section 0.6.1, we would like to construct independence tests that can be applied to more general random variables, and in particular point processes for which no natural order relation exists. Thus we turned to methods which apply in more general cases and that are based on bootstrap or permutation approaches.

Permutation approach: a pioneer concept

Sometimes called randomization tests, permutation tests have been introduced by Fisher [54] in 1935 to test the claim of a lady declaring she was able to discriminate whether the milk or the tea was first poured into the cup. Fisher constructs an experiment allowing to decide whether the null hypothesis, that is ($\mathcal{H}_0$) "the lady has no sensory discrimination", is significant or not. Based on a random permutation of several cups in which the milk was poured either before or after the tea, he computes all the probabilities under the null hypothesis of all different kinds of mistakes the lady could commit by answering only by chance. He can thus construct a valid test controlling the probability of wrongly rejecting ($\mathcal{H}_0$), that is trusting the lady's claim whereas she is in fact lying. The main argument validating this approach is that under the null hypothesis, the cups are "exchangeable" in the sense that their taste is the same. This idea is formalized and generalized below.

At the basis of permutation/randomization tests is the following fundamental assumption ($\mathcal{A}$), stated by Hoeffding's in his very nice article [84]:

($\mathcal{A}$) The set of probabilities satisfying the null hypothesis is invariant under a transformation group G.

This means that if a sample $\mathbb{Y}_n = (Y_1, \dots, Y_n)$ is drawn from a distribution satisfying the null hypothesis (e.g., the m first components $Y_1, \dots, Y_m$ have common distribution, and so do the $(n-m)$ last ones $Y_{m+1}, \dots, Y_n$), then, applying any transformation of the group G (e.g., the $m!(n-m)!$ permutations of the first m or the last $(n-m)$ components) does not affect its distribution. In other words, for all transformation g in G, $g\mathbb{Y}_n$ has the same distribution as $\mathbb{Y}_n$. In particular, it provides a simple way of mimicking the distribution of the test statistic under the null hypothesis and therefore to construct valid critical values. Notice that, in general, if the sample $\mathbb{Y}_n$ does not satisfy the null hypothesis, the previous statements may not hold, and the distribution of the transformed sample may be out of control. Yet, in the particular case of testing independence, when well applied, this approach provides a way of recreating the distribution under the null hypothesis even if the observation does not satisfy the null hypothesis. In particular, this is what happens for the permutation tests constructed in the present thesis.

Since Fisher's early work, permutation tests have been largely developed in the literature in

many testing frameworks, notably by Pitman in his series of articles [139, 140, 141]. The interested reader could refer to the books of Pesarin and Salmaso [135], Efron and Tibshirani [51] or Good [63] for reviews on such testing methods and [4, 36, 88, 100, 111, 112] for more recent works. In the particular independence testing framework, in the line of all rank tests cited above, are the ones of Scheffe [161], or Hoeffding [84]. Let us also cite the Kolmogorov-Smirnov type permutation tests of Romano [155] or van der Vaart and Wellner [173]. More recently, Gretton and his co-authors also applied the permutation approach in order to approximate the distribution of their reproducing kernel-based test statistic under the null hypothesis, as in [65, 66].
The meticulous reader may notice a gap in the literature between the 1950's and the 1990's. One of the main reasons why the permutation approach was left aside during this period is due to its high computational cost. Yet nowadays, it regains popularity thanks to huge improvements in this domain on the one hand, and to the theoretical justification of the use of Monte Carlo approximations (see, for instance, Romano and Wolf's lemma [156, Lemma 1]) on the other hand.

From Randomization to Bootstrap

Usually, permutation approaches are based on the conditional probability given the observation $\mathbb{Y}_n$, avoiding in this way the necessity of making restrictive assumptions on its distribution. This idea of working conditionally on the observation, first developed for testing, has been generalized to other domains of statistics such as estimation or construction of confidence intervals. In line with these previous works, Quenouille [144] introduced the Jackknife method for estimating the bias of estimators without making assumptions on the underlying distribution. The bootstrap approach, introduced by Efron [50] in the late 1970's, both generalizes and enhances the Jackknife.

The main idea of Efron's bootstrap is the following. Consider a sample $\mathbb{Y}_n = (Y_1, \dots, Y_n)$ with unknown distribution P. Assume you want to estimate the sample distribution of a random variable $\bar{R}_n = R_n(\mathbb{Y}_n, P)$ depending on both the data $\mathbb{Y}_n$ and its underlying distribution P. As an example, one could consider the bias of the empirical mean, that is defined by $R_n(\mathbb{Y}_n, P) = n^{-1}\sum_{i=1}^n Y_i - \int y dP(y)$, as done in [50]. The idea of the bootstrap is to replace at all levels, the true distribution P by its empirical version $P_n = n^{-1}\sum_{i=1}^n \delta_{Y_i}$, where δ_y denotes the delta measure concentrated at y. Concretely, first consider the bootstrap sample $\mathbb{Y}_n^*$ consisting of n independent and identically distributed (i.i.d.) copies drawn from the empirical distribution P_n that are obtained by picking with replacement in the original sample $\mathbb{Y}_n$. One can then estimate the distribution of $\bar{R}_n$ by the conditional distribution of $R_n^* = R_n(\mathbb{Y}_n^*, P_n)$ given $\mathbb{Y}_n$. Notice that this conditional distribution is completely known, once the sample $\mathbb{Y}_n$ is given, and takes at most n^n values obtained from resampling with replacement.

Efron's heuristics consists in saying that the conditional distribution of R_n^* given $\mathbb{Y}_n$ is close to the one of $\bar{R}_n$. The main difficulty remains in its justification. As for the permutation, this method has been vastly investigated in the literature for many kinds of statistics (see Efron and Tibshirani's [51] book for a great review). Many bootstrap type results proving that the conditional distribution of R_n^* given $\mathbb{Y}_n$ converges almost surely or in probability to the asymptotic distribution of $\bar{R}_n$ have been obtained, not only for the bootstrap of the mean as described above [20], but also, for instance, by Arcones and Giné [5] for U-statistics or by Bickel and Freedman [20] for von Mises functionals. Moreover, one should keep in mind that all bootstrap approaches that have been theoretically proved to work in the literature are based on centered statistics (meaning here that $\mathbb{E}\left[\bar{R}_n\right] = 0$). This centering necessity is

illustrated in Chapter 2.

Efron's approach has been generalized since then into many derived methods such as the weighted bootstrap. Consider, with the same notation as above, for all i in $\{1,\dots,n\}$, the random variable $M_{n,i}=\sum_{j=1}^{n}\mathbb{1}_{Y_j^*=Y_i}$ counting the number of times the ith variable Y_i appears in the bootstrap sample $\mathbb{Y}_n^*$. Then the empirical measure of the bootstrap sample can be expressed in function of the original sample by writing $P_n^*=n^{-1}\sum_{i=1}^{n}\delta_{Y_i^*}=n^{-1}\sum_{i=1}^{n}M_{n,i}\delta_{Y_i}$. The idea of the weighted bootstrap, for which Præstgaard and Wellner [143] obtained the first general results, is to replace the $M_{n,i}$'s by exchangeable weights $W_{n,i}$. For instance, one can consider deterministic non-negative weights $w_{n,i}$ such that $\sum_{i=1}^{n}w_{n,i}=n$, and a uniformly distributed random permutation Π_n in the set $\mathfrak{S}_n$ consisted of all the permutations of $\{1,\dots,n\}$. The new weights are defined by $W_{n,i}=w_{n,\Pi_n(i)}$ and the associated empirical measure is $P_n^W=n^{-1}\sum_{i=1}^{n}w_{n,\Pi_n(i)}\delta_{Y_i}$. In particular, by taking $w_{n,i}$ equal to 0 if $i=1$, and $n/(n-1)$ if $i=2,\dots,n$, one recovers the idea of the Jackknife. Other choices of weights provide other types of bootstrap such as Bretagnolle's [30] m out of n bootstrap or the wild bootstrap based on independent weights. Other aspects of the bootstrap can be found in Giné's [62] very nice review.

Many independence tests based on bootstrap approaches have been studied as, for instance, the Kolmogorov-Smirnov type ones of Romano [154] or van der Vaart and Wellner [173] (recalled in Section 0.4.3).

Notice that permutation sometimes appears in the literature as a particular case of bootstrap since both consist in resampling methods. Yet, permuting consists in resampling *without* replacement whereas bootstrapping consists in resampling *with* replacement, and in all the following, we always distinguish both approaches. Moreover, they are very different in their essence and their studies are based on completely different approaches as it may be seen in Chapter 1. What is important to keep in mind is that both permutation and bootstrap approaches work conditionally on the data and are different from other standard testing procedures in the sense that the critical values are random since depending on the data. Moreover, since for n becoming large ($n\geq 15$ for instance), the computation of all values of either the permutation or the bootstrap sample is huge (respectively at least $n!$ and n^n), Monte Carlo methods are usually applied.

The non-parametric tests presented and studied in this thesis are mainly inspired by both permutation and bootstrap tests of independence of either van der Vaart and Wellner [173] or Romano [155], himself inspired by Hoeffding's [84] permutation one. The initial motivation of this work was to adapt and generalize their test statistics to treat the neuroscience issue introduced in Section 0.6.1.

0.4.2 The mathematical testing setting

Throughout this thesis, the following notation is adopted. Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space, and $(\mathcal{X},d_{\mathcal{X}})$ be a metric space. Consider $X=(X^1,X^2)$ a random variable on $\mathcal{X}^2$, that is a measurable function from $(\Omega,\mathcal{A})$ to $(\mathcal{X}^2,\mathcal{B}_{\mathcal{X}^2})$, where $\mathcal{B}_{\mathcal{X}^2}$ denotes the σ-algebra of Borel sets of $(\mathcal{X}^2,d_{\mathcal{X}^2})$ and $d_{\mathcal{X}^2}$ is a product metric. The random variable X is said to have distribution P and marginals P^1 and P^2, if for all Borel set B of $\mathcal{X}^2$, $\mathbb{P}(X\in B)=P(B)$ and for all Borel sets B^1 and B^2 of $\mathcal{X}$, $\mathbb{P}\big(X^1\in B^1\big)=P^1(B^1)$ and $\mathbb{P}\big(X^2\in B^2\big)=P^2(B^2)$. The product of the marginals $P^1\otimes P^2$ is then defined for all Borel sets B^1 and B^2 of $\mathcal{X}$ by $P^1\otimes P^2(B^1\times B^2)=P^1(B^1)P^2(B^2)$, and the variables X^1 and X^2 are independent if and only if $P=P^1\otimes P^2$.

Two different data frameworks. The first data framework, mainly considered in Chapters 1 and 2, is a point process framework adapted to the biological application in neuroscience described in Section 0.6. In this case, $\mathcal{X}$ is the set of all countable sets of points in $[0,1]$, equipped with a metric issued from the Skorohod topology as detailed in Appendix A.2.2.
The second data framework, mainly considered in Chapter 4, is the standard density framework where $\mathcal{X}$ is a compact real interval, which is considered here as $[0,1]$ without loss of generality, and the random variables on $\mathcal{X}^2$ are assumed to have a density in $\mathbb{L}_2(\mathcal{X}^2)$ with respect to the Lebesgue measure.

The independence testing issue. In all the sequel, $\mathbb{X}_n = (X_1, \dots, X_n)$ denotes a sample of n i.i.d. copies $X_i = (X_i^1, X_i^2)$ of a random variable $X = (X^1, X^2)$ on $\mathcal{X}^2$, with distribution P and marginals P^1 and P^2. It will be referred to as observation, (observed) sample or original data throughout this thesis. Given the observation of the sample $\mathbb{X}_n$, the aim is to test the null hypothesis $(\mathcal{H}_0)$ "the coordinates X^1 and X^2 are independent" against the alternative $(\mathcal{H}_1)$ "they are not", that is equivalent to test

$$(\mathcal{H}_0)\ P = P^1 \otimes P^2 \quad \text{against} \quad (\mathcal{H}_1)\ P \neq P^1 \otimes P^1.$$

Throughout this thesis, a statistical test of $(\mathcal{H}_0)$ against $(\mathcal{H}_1)$ is a function Δ of the observation $\mathbb{X}_n$, possibly depending on n, with values in $\{0,1\}$, $\Delta(\mathbb{X}_n) = 1$ meaning the null hypothesis $(\mathcal{H}_0)$ is rejected in favor of the alternative $(\mathcal{H}_1)$, $\Delta(\mathbb{X}_n) = 0$ meaning $(\mathcal{H}_0)$ cannot be rejected in view of the data, and in that sense, is accepted. A statistical test has two different kinds of errors. The *first kind error*, also referred to as a false positive, consists in wrongly rejecting the null hypothesis $(\mathcal{H}_0)$, and the *second kind error*, also called a false negative, consists in wrongly accepting $(\mathcal{H}_0)$. Denote by $\mathcal{P}_0$ (respectively $\mathcal{P}_1$) the sets of all probability measures on $(\mathcal{X}^2, \mathcal{B}_{\mathcal{X}^2})$ satisfying $(\mathcal{H}_0)$ (respectively $(\mathcal{H}_1)$), that is the set of probability measures equal to (respectively different from) the product of their marginals. One can then define respectively the *first and second kind error rates* by

$$\sup_{P \in \mathcal{P}_0} \mathbb{P}_P(\Delta(\mathbb{X}_n) = 1) \quad \text{and} \quad \sup_{P \in \mathcal{P}_1} \mathbb{P}_P(\Delta(\mathbb{X}_n) = 0)\,,$$

where the index P in $\mathbb{P}_P(\Delta(\mathbb{X}_n) = 0)$ means that $\mathbb{X}_n$ is a sample of i.i.d. random variables with distribution P.
For any prescribed α in $(0,1)$, a test Δ is said to be *exactly of level* α if its first kind error rate is controlled by α, and this for any sample size n. It is moreover said to be *asymptotically of size* α if its first kind error rate converges to α as the sample size n tends to $+\infty$.
The *power* of the test is defined by the function $P \in \mathcal{P}_1 \mapsto 1 - \mathbb{P}_P(\Delta(\mathbb{X}_n) = 0)$. Finally, a statistical test is said to be *consistent* against an alternative P in $\mathcal{P}_1$ if its power, under this alternative, converges to one as the sample size n tends to $+\infty$.

Construction of a test. Given the observation of a sample $\mathbb{X}_n$ of i.i.d. random variables with unknown distribution P, there are several ways of constructing a statistical test of prescribed level α in $(0,1)$. All methods begin with a real-valued *test statistic* T, that is a function of $\mathbb{X}_n$, well adapted to the testing problem. The knowledge of its distribution under the null hypothesis, that is when P satisfies $(\mathcal{H}_0)$, also called the *null distribution* in the sequel, is the central goal in the construction of a test.

Then, the standard approach consists in defining a *critical region* R_α (depending on α) included in $\mathbb{R}$. The test is then defined by $\Delta_\alpha(\mathbb{X}_n) = \mathbb{1}_{T(\mathbb{X}_n) \in R_\alpha}$. In general, the critical region

R_α is established such that the test is either exactly of prescribed level α (if possible) or asymptotically of size α. One may then distinguish three particular kinds of tests that are *upper-tailed*, *lower-tailed* or *two-tailed* tests where the critical regions are respectively of the forms $(c_\alpha^+, +\infty)$, $(-\infty, c_\alpha^-)$ and $(-\infty, c_{\alpha/2}^-) \cup (c_{\alpha/2}^+, +\infty)$. Moreover, if the null distribution, or its asymptotic limit is known, the critical values c_α^+ and c_α^- are chosen to be the quantiles of order $(1-\alpha)$ and α of this distribution.

Another approach is based on the notion of *p-value* (also called *achieved significance level* by Efron and Tibshirani [51]). The p-value corresponding to a testing procedure $\boldsymbol{\Delta} = \{\Delta_\alpha\}_{\alpha\in(0,1)}$ as above is defined by $\hat{\alpha}(\mathbb{X}_n) = \sup\{\alpha'; \Delta_{\alpha'}(\mathbb{X}_n) = 0\}$. It is the greatest level of significance α' under which the test $\Delta_{\alpha'}$ accepts the null hypothesis in view of the observation $\mathbb{X}_n$. Naturally, the test can usually be viewed as $\Delta_\alpha(\mathbb{X}_n) = \mathbb{1}_{\hat{\alpha}(\mathbb{X}_n)\leq\alpha}$. Such a test is exactly of any level if for all u in $(0,1)$, $\mathbb{P}(\hat{\alpha}(\mathbb{X}_n) \leq u) \leq u$ under the null hypothesis. Informally, the p-value represents the reliability of the null hypothesis given the observation. The smaller it is, the stronger the evidence against $(\mathcal{H}_0)$ is. In particular, as mentioned by Efron and Tibshirani in [51], one commonly admits that a p-value $\hat{\alpha}(\mathbb{X}_n)$ smaller than 0.1, 0.05, 0.025 and 0.01 respectively correspond to "borderline", "reasonably strong", "strong" and "very strong" evidence against $(\mathcal{H}_0)$. When the null distribution, or the asymptotic null distribution of $T(\mathbb{X}_n)$ is known, the p-value can also be expressed in terms of the left-continuous version of the corresponding cumulative distribution function F_0^-. For instance, the p-value corresponding to the upper-tailed test, is equal to $\hat{\alpha}(\mathbb{X}_n) = 1 - F_0^-(T(\mathbb{X}_n))$. In particular, it is exactly the probability under the null hypothesis of observing a test statistic greater than or equal to the observed value $T(\mathbb{X}_n)$, or more precisely, $\mathbb{P}\big(T(\mathbb{X}_n^0) \geq T(\mathbb{X}_n)|\mathbb{X}_n\big)$ where $\mathbb{X}_n^0$ is an i.i.d. sample under the null hypothesis and independent of $\mathbb{X}_n$.

In this thesis, both constructions are considered (see Section 0.4.3): critical values are provided in Chapter 1 and are translated into p-values in Chapter 2.

Optimality in the minimax sense. Usually, there are many possible testing procedures for a particular testing problem. A natural question is to know how to compare two tests, and before that, how to describe the performance of a test. In this thesis, the non-asymptotic performance of a test is expressed in terms of uniform separation rates as described here.

Given two prescribed levels of error rates α and β in $(0,1)$, the aim is to construct a test of $(\mathcal{H}_0)$ "$P \in \mathcal{P}_0$" which is exactly of level α, and for which the second kind error rate is controlled by β. Usually, the critical region is adjusted such that the first kind error rate is automatically controlled by α. Since no variability is left to force the test to control the second kind error rate, the idea is to restrain the set of all alternatives $\mathcal{P}\backslash\mathcal{P}_0$ on which the supremum in the second kind error rate is taken.

Intuitively, the further the alternative is from the null hypothesis, the more able the test is to reject it. The idea is thus to introduce the set of alternatives at distance at least $\rho > 0$ from the null hypothesis. Yet, since this subset of alternatives is often still too large, it is in addition restrained to a family $\{\mathcal{Q}_\nu\}_{\nu\in\mathcal{M}}$ of regularity subspaces of $\mathcal{P}$. For instance, as in the density framework of Chapter 4, each $\mathcal{Q}_\nu$ may denote a subspace of the set of probability measures with density functions admitting some smoothness properties with parameter ν.

Definition 0.4.1. The *uniform separation rate* $\rho(\Delta_\alpha, \mathcal{Q}_\nu, \beta)$ of a test Δ_α which is exactly of level α, over a regularity class of distributions $\mathcal{Q}_\nu$ with respect to some distance $\bar{d}$ over the set of probability measures on $\mathcal{X}^2$ is defined by

$$\rho(\Delta_\alpha, \mathcal{Q}_\nu, \beta) = \inf\left\{\rho > 0\ ;\ \sup_{P\in\mathcal{Q}_\nu;\ \bar{d}(P,\mathcal{P}_0)>\rho} \mathbb{P}_P(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta\right\}.$$

Intuitively, it is the smallest distance from the null hypothesis for which the test well detects the alternatives of regularity ν with a second kind error rate controlled by β.
Hence, when comparing tests of $(\mathcal{H}_0)$ "$P \in \mathcal{P}_0$" which are all exactly of level α, the one with the smallest uniform separation rate is the most efficient one since it is capable of detecting more alternatives than the other ones. The natural idea to describe the non-asymptotic performance of a test is thus to compare it with the "best" one in terms of uniform separation rates. To do so, the *minimax separation rate of testing* $(\mathcal{H}_0)$ "$P \in \mathcal{P}_0$", has been introduced by Baraud [13], and is defined by

$$\rho(\mathcal{Q}_\nu, \alpha, \beta) = \inf\left\{\rho(\Delta_\alpha, \mathcal{Q}_\nu, \beta)\right\},$$

where the infimum is taken over all tests Δ_α of $(\mathcal{H}_0)$ which are exactly of level α. A test exactly of level α which achieves, up to a multiplicative constant, the minimax separation rate is said to be *optimal in the minimax sense*.

The notion of optimality in the minimax sense was first introduced by Ingster in a fundamental series of papers [91, 95] but from an asymptotic point of view. The (asymptotic) *minimax rate of testing* $\rho_n(\mathcal{Q}_\nu)$ over the regularity space $\mathcal{Q}_\nu$ satisfies an upper-bound condition that is

$$\forall \alpha, \beta > 0,\ \exists C > 0 \text{ and a test } \Delta^* \text{ such that} \quad \begin{cases} \limsup\limits_{n\to+\infty} \sup\limits_{P\in\mathcal{P}_0} \mathbb{P}_P(\Delta^*(\mathbb{X}_n) = 1) \leq \alpha, \\ \limsup\limits_{n\to+\infty} \sup\limits_{P\in\mathcal{Q}_\nu,\ \bar{d}(P,\mathcal{P}_0)>C\rho_n(\mathcal{Q}_\nu)} \mathbb{P}_P(\Delta^*(\mathbb{X}_n) = 0) \leq \beta. \end{cases} \tag{0.4.1}$$

and a lower-bound condition that is for all rate ρ'_n such that $\rho'_n/\rho_n(\mathcal{Q}_\nu) \underset{n\to+\infty}{\longrightarrow} 0$,

$$\inf_{\Delta} \left\{ \sup_{P\in\mathcal{P}_0} \mathbb{P}_P(\Delta(\mathbb{X}_n) = 1) + \sup_{P\in\mathcal{Q}_\nu,\ \bar{d}(P,\mathcal{P}_0)>\rho'_n} \mathbb{P}_P(\Delta(\mathbb{X}_n) = 0) \right\} \underset{n\to+\infty}{\longrightarrow} 1.$$

Intuitively, the upper-bound condition ensures the minimax rate of testing is achieved up to a constant, and the lower-bound condition ensures it is the optimal one. In the literature, the *minimax separation rate of testing* is often equivalent to the (asymptotic) *minimax rate of testing*. Yet, since in the following, non-asymptotic results are expected, we will focus on the first one.

If in addition of being optimal in the minimax sense on $\mathcal{Q}_\nu$, a testing procedure does not depend on the regularity parameter ν, the test is said to be *adaptive in the minimax sense*. A cost in $\ln(n)$ or $\ln(\ln(n))$ is sometimes inevitable for adaptivity depending on the class $\mathcal{Q}_\nu$ of alternatives (see Spokoiny [167]).

0.4.3 Construction of two non-parametric tests of independence

For a matter of simplicity, only the upper-tailed tests are presented in this introduction, but the lower-tailed and the two-tailed tests are also defined and studied in Chapter 1.
Assume we observe a sample $\mathbb{X}_n$ of n i.i.d. random variables with distribution P on $\mathcal{X}^2$, and fix a prescribed level α in $(0, 1)$.

The Test Statistic

Before describing how both bootstrap and permutation approaches are applied to construct critical values, let us introduce the test statistic. It is defined by $\sqrt{n}U_{n,h}(\mathbb{X}_n)$, where $U_{n,h}(\mathbb{X}_n)$

is a U-statistic based on a measurable symmetric kernel $h : \mathcal{X}^2 \times \mathcal{X}^2 \to \mathbb{R}$, that is

$$U_{n,h}(\mathbb{X}_n) = \frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} h(X_i, X_j). \tag{0.4.2}$$

The kernel is assumed to satisfy the following centering assumption:

$$\int_{\mathcal{X}^2 \times \mathcal{X}^2} h(x,y) dP(x) dP(y) = 0,$$

which guarantees that the test statistic $\sqrt{n} U_{n,h}(\mathbb{X}_n)$ is centered under the null hypothesis. Notice that this centering assumption is essential when applying bootstrap approaches as illustrated in Chapter 2, Section 2.3. Generally, U-statistics provide great tools for distribution free testing since they are reliable non-parametric estimators and their asymptotic behavior has been largely studied during the last decades (see, for instance, [164, Chapter 5]). A brief introduction on U-statistics is made in Appendix A.3.

An important particular case considered in this thesis, called the *Linear case*, is when the kernel h is of the form h_φ defined for all $x = (x^1, x^2)$ and $y = (y^1, y^2)$ in $\mathcal{X}^2$ by

$$h_\varphi(x,y) = \frac{1}{2} \left[\varphi(x^1, x^2) + \varphi(y^1, y^2) - \varphi(x^1, y^2) - \varphi(y^1, x^2) \right], \tag{0.4.3}$$

and $\varphi : \mathcal{X}^2 \to \mathbb{R}$ denotes a measurable function. Then, the centering assumption is automatically satisfied. Notice that in this case, the U-statistic is an unbiased estimator of

$$\int_{\mathcal{X}^2} \varphi(x^1, x^2) \left[dP(x^1, x^2) - dP^1(x^1) dP^2(x^2) \right], \tag{0.4.4}$$

and this without any assumption on the underlying distribution P of $\mathbb{X}_n$. In the literature, the quantity in (0.4.4) is a starting point of many independence tests since, for well-chosen φ, or possibly a supremum over several ones, it provides a pseudo-distance between the joint distribution P and the product of its marginals $P^1 \otimes P^2$ and thus its equality to zero characterizes independence.

When considered with a supremum, one recovers the Kolmogorov-Smirnov type test statistics of Romano [154, 155] or van der Vaart and Wellner [173], which can be written with our notation as

$$H_n(\mathbb{X}_n) = \frac{(n-1)}{n} \times \sup_{(v^1, v^2) \in \mathcal{V}^1 \times \mathcal{V}^2} \left| \sqrt{n} U_{n, h_{\varphi_{(v^1, v^2)}}}(\mathbb{X}_n) \right|, \tag{0.4.5}$$

where respectively

- $\mathcal{V}^1$ and $\mathcal{V}^2$ are countable Vapnik-Chervonenkis classes of subsets of $\mathcal{X}$, and $\varphi_{(v^1,v^2)}(x^1, x^2) = \mathbb{1}_{v^1}(x^1) \mathbb{1}_{v^2}(x^2)$,
- $\mathcal{V}^1$ and $\mathcal{V}^2$ are well-chosen classes of measurable real-valued functions on $\mathcal{X}$, and $\varphi_{(v^1,v^2)}(x^1, x^2) = v^1(x^1) v^2(x^2)$.

One may notice that only product type functions φ are considered in these previous works and, to our knowledge, in the literature. Yet, because of the neurobiological motivation, more general forms of φ need to be studied, such as the delayed coincidence count φ_δ^{coinc} introduced in Section 0.6.1. This generalization required further developments. It is one of the main statistical contributions of this thesis.

A bootstrap test of independence

Inspired by Romano's [154] and van der Vaart's [173] independence tests, we first consider a bootstrap approach for constructing a non-parametric independence test. Instead of considering Efron's "naive" approach, which consists in resampling with replacement in the original sample of couples $(X_1, \dots, X_n)$, we follow the idea of Romano [154] and van der Vaart and Wellner [173] which is to sample from the product of the empirical marginal distributions, forcing in this way the independence between the coordinates in the bootstrap sample. More precisely, consider the empirical marginals denoted for $j = 1, 2$ by

$$P_n^j = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i^j}.$$

The bootstrap sample, denoted by $\mathbb{X}_n^* = (X_1^*, \dots, X_n^*)$, is an i.i.d. sample with distribution $P_n^1 \otimes P_n^2$. Concretely it consists in picking with replacement in the first coordinates $\{X_1^1, \dots X_n^1\}$ and independently, picking with replacement in the second coordinates $\{X_1^2, \dots, X_n^2\}$. It is then clear that conditionally on the sample $\mathbb{X}_n$, the coordinates of each X_i^* are independent, and thus the bootstrap sample $\mathbb{X}_n^*$ satisfies the independence hypothesis $(\mathcal{H}_0)$. Hence, it provides a simple way of reconstructing the null distribution.

According to the bootstrap paradigm, the conditional distribution of the bootstrap test statistic $\sqrt{n}U_{n,h}(\mathbb{X}_n^*)$ given $\mathbb{X}_n$ should be close to the null distribution. The critical value $q_{1-\alpha,n}^*(\mathbb{X}_n)$ is thus defined by the $(1-\alpha)$-quantile of the conditional distribution of $\sqrt{n}U_{n,h}(\mathbb{X}_n^*)$ given $\mathbb{X}_n$, and the *upper-tailed bootstrap test* rejects independence when

$$\sqrt{n}U_{n,h}(\mathbb{X}_n) > q_{1-\alpha,n}^*(\mathbb{X}_n).$$

Notice that the critical value is random, since it depends on $\mathbb{X}_n$, and that it may be exactly computed by considering the n^{2n} possible bootstrap samples. Indeed, let

$$\sqrt{n}U_{n,h}^{*(1)}(\mathbb{X}_n) \ \leq \ \dots \ \leq \ \sqrt{n}U_{n,h}^{*(n^{2n})}(\mathbb{X}_n)$$

be the ordered values that the bootstrap test statistic can take given $\mathbb{X}_n$. Then, the critical value satisfies $q_{1-\alpha,n}^*(\mathbb{X}_n) = \sqrt{n}U_{n,h}^{*(\lceil n^{2n}(1-\alpha)\rceil)}(\mathbb{X}_n)$.

As already mentioned, even for moderately large values of n, the computation of the exact conditional quantile $q_{1-\alpha,n}^*(\mathbb{X}_n)$ is very costly. Hence, the following Monte Carlo method is applied to approximate it. Compute B_n i.i.d. bootstrap samples (conditionally on $\mathbb{X}_n$), denoted by $\mathbb{X}_n^{*1}, \dots, \mathbb{X}_n^{*B_n}$. Let $U^{*b} = U_{n,h}(\mathbb{X}_n^{*b})$ for all $1 \leq b \leq B_n$ and consider the corresponding order statistic $U^{*(1)} \leq \dots \leq U^{*(B_n)}$. Then, the Monte Carlo approximation of the quantile is given by $q_{1-\alpha,n}^{*MC}(\mathbb{X}_n) = \sqrt{n}U^{*(\lceil B_n(1-\alpha)\rceil)}$, and the *upper-tailed bootstrap test with Monte Carlo approximation* rejects independence when

$$\sqrt{n}U_{n,h}(\mathbb{X}_n) > q_{1-\alpha,n}^{*MC}(\mathbb{X}_n).$$

The corresponding p-value of this test is given by

$$\frac{1}{B_n} \sum_{b=1}^{B_n} \mathbb{1}_{U^{*b} \geq U_{n,h}(\mathbb{X}_n)}.$$

Note that the *upper-tailed bootstrap test with Monte Carlo approximation* is exactly the test rejecting independence when this p-value is smaller than or equal to α. Indeed,

$$\begin{aligned}\sqrt{n}U_{n,h}(\mathbb{X}_n) > q^{*MC}_{1-\alpha,n}(\mathbb{X}_n) &\Leftrightarrow U_{n,h}(\mathbb{X}_n) > U^{*(\lceil B_n(1-\alpha)\rceil)}\\ &\Leftrightarrow \sum_{b=1}^{B_n} \mathbb{1}_{U^{*b}<U_{n,h}(\mathbb{X}_n)} \geq \lceil B_n(1-\alpha)\rceil\\ &\Leftrightarrow \sum_{b=1}^{B_n} \mathbb{1}_{U^{*b}\geq U_{n,h}(\mathbb{X}_n)} \leq \lfloor \alpha B_n\rfloor\\ &\Leftrightarrow \sum_{b=1}^{B_n} \mathbb{1}_{U^{*b}\geq U_{n,h}(\mathbb{X}_n)} \leq \alpha B_n.\end{aligned}$$

A permutation test of independence

One may notice that under the independence hypothesis ($\mathcal{H}_0$), the distribution of the sample $\mathbb{X}_n$ is invariant under all transformations that consist in permuting the second coordinates of the sample. The fundamental assumption ($\mathcal{A}$) for permutation testing is thus satisfied and such approach can then be applied for this problematic.

The permutation method, inspired by Hoeffding [84], and van der Vaart and Wellner [173], consists in the following. Let Π_n be a random permutation uniformly distributed on the set $\mathfrak{S}_n$ of all permutations of $\{1,\dots,n\}$, and independent of $\mathbb{X}_n$. Then, consider the permuted sample defined by $\mathbb{X}_n^{\Pi_n} = (X_1^{\Pi_n},\dots,X_n^{\Pi_n})$ where each $X_i^{\Pi_n} = (X_i^1, X^2_{\Pi_n(i)})$ is obtained from permuting the second coordinates according to Π_n.

As for the bootstrap test, the critical value $q^\star_{1-\alpha,n}(\mathbb{X}_n)$ is defined by the $(1-\alpha)$-quantile of the conditional distribution of the permuted test statistic $\sqrt{n}U_{n,h}(\mathbb{X}_n^{\Pi_n})$ given $\mathbb{X}_n$, and the *upper-tailed permutation test* rejects independence when

$$\sqrt{n}U_{n,h}(\mathbb{X}_n) > q^\star_{1-\alpha,n}(\mathbb{X}_n).$$

Once again, the critical value is random, depending on the observed sample $\mathbb{X}_n$, and is possible to compute exactly. Since Π_n is uniformly distributed on $\mathfrak{S}_n$ and independent of $\mathbb{X}_n$, the conditional distribution of $\sqrt{n}U_{n,h}\big(\mathbb{X}_n^{\Pi_n}\big)$ given $\mathbb{X}_n$ is discrete and takes the values $\{\sqrt{n}U_{n,h}(\mathbb{X}_n^{\pi_n})\}_{\pi_n\in\mathfrak{S}_n}$. Let

$$\sqrt{n}U^{\star(1)}_{n,h}(\mathbb{X}_n) \leq \ldots \leq \sqrt{n}U^{\star(n!)}_{n,h}(\mathbb{X}_n)$$

be the corresponding ordered values. Then, $q^\star_{1-\alpha,n}(\mathbb{X}_n) = \sqrt{n}U^{\star(\lceil n!(1-\alpha)\rceil)}_{n,h}(\mathbb{X}_n)$.

Once again, Monte Carlo methods are used in practice to approximate both the quantile and the p-value. However, for exact level considerations explained in Section 0.5.2, the method for the permutation test slightly differs from the one for the bootstrap test, and consists in adding to the collection of permuted test statistics, the original one.
More precisely, compute B_n i.i.d. uniform permutations of $\{1,\dots,n\}$, denoted by $\Pi_n^1,\dots,\Pi_n^{B_n}$. Let

$$U^{\star b} = U_{n,h}\left(\mathbb{X}_n^{\Pi_n^b}\right) \text{ for all } 1\leq b\leq B_n, \quad \text{and} \quad U^{\star B_n+1} = U_{n,h}(\mathbb{X}_n).$$

Consider the order statistic $U^{\star(1)} \leq \cdots \leq U^{\star(B_n+1)}$. Then, the approximated quantile is given by $q^{\star MC}_{1-\alpha,n}(\mathbb{X}_n) = \sqrt{n}U^{\star(\lceil(1-\alpha)(B_n+1)\rceil)}$ and the *upper-tailed permutation test with Monte Carlo approximation* rejects independence when

$$\sqrt{n}U_{n,h}(\mathbb{X}_n) > q^{\star MC}_{1-\alpha,n}(\mathbb{X}_n).$$

The corresponding p-value of this test is given by

$$\frac{1}{B_n+1}\left(1+\sum_{b=1}^{B_n}\mathbb{1}_{U^{*b}\geq U_{n,h}(\mathbb{X}_n)}\right).$$

As for the bootstrap test, one can easily check that the *upper-tailed permutation test with Monte Carlo approximation* is equivalent to the test that rejects independence when this p-value is smaller than α.

Let us point out the difference between the star $*$ corresponding to the bootstrap approach, and the star $\star$ corresponding to the permutation approach. Now that we have introduced the new testing procedures studied in this thesis, we may ask what their performance is.

0.5 How to study the performance?

The study of the performance of a test goes through the analysis of both its first and its second kind error rates. Usually, when working with bootstrap or permutation techniques, the asymptotic properties of the tests are investigated (see, for instance, [155, 173]). In this thesis, we want to go further and also study non-asymptotic properties of the permutation test which is already known to be exactly of prescribed level. Chapter 1 is devoted to the study of the asymptotic performance of the bootstrap and permutation tests in the point process framework, and Chapter 4 is devoted to the study of the non-asymptotic properties of the two-tailed permutation test in the density framework.

0.5.1 Asymptotic study

Several ways of studying the asymptotic properties of a test have been introduced in the literature. The one adopted in this thesis consists in checking separately that both kind error rates are controlled when the sample size grows to infinity, and more precisely that the test is asymptotically of prescribed size and consistent against some alternatives.

To do so, the aim when applying a bootstrap (respectively permutation) approach is to prove that the conditional distribution of the bootstrapped (respectively permuted) test statistic given the original sample converges to the asymptotic distribution of the test statistic under the null hypothesis, namely the asymptotic null distribution. If this result holds only when the original sample satisfies the null hypothesis, we say that the resampling method *mimics* the null distribution. In this case, since this result only holds under the null hypothesis, it only concerns the size property. Furthermore, if this result also holds under some alternatives, we say that the resampling method *reconstructs* the null distribution. In addition to the size properties, it then also provides the consistency of the test.

On van der Vaart and Wellner's and Romano's results. As an example, let us recall the asymptotic results obtained by van der Vaart and Wellner [173, Section 3.8] for their bootstrap and permutation tests of independence based on the Kolmogorov-Smirnov type test statistic H_n defined in (0.4.5).

- First, they prove that, if the sets $\mathcal{V}^1$ and $\mathcal{V}^2$ are Donsker classes (informally characterized by the existence of a central limit theorem that holds uniformly over these classes, see [173, p.81]), and if $\mathbb{X}_n$ is an i.i.d. sample from a distribution satisfying $(\mathcal{H}_0)$, that is $P = P^1 \otimes P^2$, then $H_n(\mathbb{X}_n)$ converges in distribution to a limit distribution $\mathcal{L}_{P^1 \otimes P^2}$ depending on P^1 and P^2.
- On the one hand, for the bootstrap approach, they prove that, whether P does or does not satisfy the null hypothesis, the conditional distribution of the bootstrapped test statistic given $\mathbb{X}_n$ weakly converges to the same limit $\mathcal{L}_{P^1 \otimes P^2}$, and this, almost surely in the original sample $\mathbb{X}_n$. They thus prove that the bootstrap approach reconstructs the null distribution.
- On the other hand, the reconstruction ability of the permutation approach is left as an open question, although they believe the permuted test statistics have a similar asymptotic behavior than the bootstrapped test statistics.

This similarity between bootstrap and permutation is proved by Romano [155, Proposition 3.1] in the particular case of indicator functions over sets in a Vapnik-Chervonenkis class. More precisely, Romano proves that the infinity norm between both cumulative distribution functions of the conditional distributions of his bootstrapped and his permuted test statistics given $\mathbb{X}_n$ converges in probability to zero, and this whether P satisfies or not the null hypothesis (since both bootstrapped and permuted sample are forced to satisfy $(\mathcal{H}_0)$). To do so, he first proves (in [154]) similar results as van der Vaart and Wellner's ones about the null distribution and the bootstrap approach. As above, denote by $\mathcal{L}_{P^1 \otimes P^2}$ the shared limit distribution. Then, based on Hoeffding's result [84, Theorem 3.2], he proves that the conditional distribution of the permuted test statistic weakly converges in probability to the same limit $\mathcal{L}_{P^1 \otimes P^2}$ and this whether the original sample $\mathbb{X}_n$ satisfies or not $(\mathcal{H}_0)$. He thus proves that both bootstrap and permutation approaches reconstruct the null distribution. Notice however that, to prove reconstruction under the alternative, he theoretically uses a permutation approach which is not exactly the one described above in the sense that he does not consider permutations with fixed points, as far as we understand his argument.

Contribution of this thesis

Following the same outline, we achieved the ensuing results in the point process framework. Once again, only the study of the upper-tailed test is illustrated here for a matter of simplicity, but the results for the lower-tailed and the two-tailed tests are also stated in Chapter 1. Instead of looking at the cumulative distribution functions as did Romano, all the convergences are expressed in Chapter 1 in terms of the $\mathbb{L}_2$-Wasserstein metric (which has the main particularity that a convergence in such distance implies a weak convergence, see Appendix A.1.2, Proposition A.1.2) as it is commonly done when studying bootstrap approaches (see, for instance, [20]).

Convergence results. In the context of this thesis, the standard central limit theorem for non-degenerate U-statistics (see Appendix A.3 for the standard result, or Chapter 1,

Proposition 1.3.5 for its version in the Wasserstein metric d_2) states that, under mild moment and non-degeneracy assumptions, if $P = P^1 \otimes P^2$, then the true distribution of the test statistic $\sqrt{n}U_{n,h}(\mathbb{X}_n)$, denoted by $\mathcal{L}\left(\sqrt{n}U_{n,h}, P\right)$, converges in the Wasserstein metric to a centered Gaussian distribution with variance $\sigma^2_{P^1\otimes P^2}$ depending on P, that is exactly

$$d_2\left(\mathcal{L}\left(\sqrt{n}U_{n,h}, P^1 \otimes P^2\right), \mathcal{N}\left(0, \sigma^2_{P^1\otimes P^2}\right)\right) \underset{n\to+\infty}{\longrightarrow} 0. \tag{0.5.1}$$

As a consequence, the null distribution weakly converges to $\mathcal{N}(0, \sigma^2_{P^1\otimes P^2})$.

Moreover, both bootstrap and permutation approaches are shown to reconstruct the null distribution. For the bootstrap test, the distance between the conditional distribution of the bootstrapped test statistic given $\mathbb{X}_n$, denoted $\mathcal{L}\left(\sqrt{n}U_{n,h}, P_n^1 \otimes P_n^2 \middle| \mathbb{X}_n\right)$, and the null distribution directly converges to zero, and this without using the Gaussian asymptotic limit as an intermediary. More precisely, in Theorem 1.3.1, we prove that

$$d_2\left(\mathcal{L}\left(\sqrt{n}U_{n,h}, P_n^1 \otimes P_n^2 \middle| \mathbb{X}_n\right), \mathcal{L}\left(\sqrt{n}U_{n,h}, P^1 \otimes P^2\right)\right) \underset{n\to+\infty}{\longrightarrow} 0, \ P\text{-a.s. in } (X_i)_i. \tag{0.5.2}$$

The arguments in the proof are standard (see [20, 41, 118]), based on common tools such as the weak convergence of empirical measures in separable spaces [174, Theorem 3] (see Appendix A.1.2, Theorem A.1.4), and Skorohod's representation Theorem [46, Theorem 11.7.2] (see Appendix A.1.2, Theorem A.1.2). The main difficulties encountered were essentially due to the nature of our random variables, namely point processes.

Concerning the permutation approach, we had to restrict our study to the *Linear case* (which corresponds to the case when the kernel of the U-statistic h is of the form h_φ as in (0.4.3)) for technical reasons. In this case, the permuted test statistic can be written as a normalized centered permuted sum, that is

$$\begin{aligned}
\sqrt{n}U_{n,h_\varphi}\left(\mathbb{X}_n^{\Pi_n}\right) &= \frac{\sqrt{n}}{n-1}\left(\sum_{i=1}^n \varphi\left(X_i^1, X^2_{\Pi_n(i)}\right) - \frac{1}{n}\sum_{i,j}\varphi\left(X_i^1, X_j^2\right)\right) \\
&= \frac{\sqrt{n}}{n-1}\left(\sum_{i=1}^n \varphi\left(X_i^1, X^2_{\Pi_n(i)}\right) - \mathbb{E}\left[\sum_{i=1}^n \varphi\left(X_i^1, X^2_{\Pi_n(i)}\right)\middle| \mathbb{X}_n\right]\right).
\end{aligned} \tag{0.5.3}$$

One of the main difficulties encountered for this approach comes from the fact that the chosen function φ is not assumed to be of the form of a product anymore, as in Romano's [155] or van der Vaart and Wellner's [173] tests. In order to validate the consistency of this approach, we proved a new combinatorial central limit theorem for permuted sums (see Chapter 1, Theorem 1.4.1). More precisely we prove that the distance between the conditional distribution of the permuted test statistic given $\mathbb{X}_n$, denoted by $\mathcal{L}\left(\sqrt{n}U_{n,h_\varphi}, P_n^\star \middle| \mathbb{X}_n\right)$ (where $P_n^\star$ stands for the conditional distribution of $\mathbb{X}_n^{\Pi_n}$ given $\mathbb{X}_n$), and the asymptotic null distribution obtained in (0.5.1) tends to zero, that is,

$$d_2\left(\mathcal{L}\left(\sqrt{n}U_{n,h_\varphi}, P_n^\star \middle| \mathbb{X}_n\right), \mathcal{N}\left(0, \sigma^2_{P^1\otimes P^2}\right)\right) \underset{n\to+\infty}{\longrightarrow} 0, \ \text{in probability.} \tag{0.5.4}$$

Notice that many combinatorial central limit theorems have been shown in the literature. The original one obtained by Wald and Wolfowitz [178] for randomly permuted deterministic sums of the product type, such as $\sum_{i=1}^n b_i \times c_{\Pi_n(i)}$, has been largely generalized: for instance, not

necessarily product type terms are considered by Hoeffding [83], Lindeberg-type assumptions are used by Motoo [127], random terms are considered by Dwass [49], and more recently, weighted U-statistics are considered by Shapiro and Hubert [165]. More details can be found in the introduction of Chapter 3. The main novelty of our result is that it is free from the restrictive and usual exchangeability assumption on the permuted random variables and thus holds under any alternative. This result notably provides that the permutation approach reconstructs (and not only mimics) the null distribution, and thus partly (as we did not consider the supremum case) answers the open question left by van der Vaart and Wellner in [173]. It is one of the newest results presented here, and its scope is well beyond the only generalization to the point processes setting, since the proof holds for any kind of random variable with values in a separable space.

Asymptotic properties of the tests. Finally, we deduced from these convergence results that both bootstrap and permutation tests are asymptotically of prescribed size and consistent against any reasonable alternative P for which the expected value of the test statistic under P, namely $\int_{\mathcal{X}^2\times\mathcal{X}^2} h(x,y)dP(x)dP(y)$, is strictly positive (recall that it is equal to zero under the null hypothesis because of the centering assumption).
The proofs are standard, and heuristically described for the bootstrap (respectively permutation) test below.
On the one hand, the asymptotic size is obtained as follows. Assume that P satisfies $(\mathcal{H}_0)$.

- First, thanks to (0.5.1), one obtains that the test statistic $\sqrt{n}U_{n,h}(\mathbb{X}_n)$ converges in distribution to a random variable with distribution $\mathcal{N}\left(0,\sigma^2_{P^1\otimes P^2}\right)$, say $Z_{P^1\otimes P^2}$.

- Then, since the limit distribution is continuous, combining both (0.5.1) and (0.5.2) (respectively thanks to (0.5.4)), one can deduce that the critical value $q^*_{1-\alpha,n}(\mathbb{X}_n)$ (respectively $q^\star_{1-\alpha,n}(\mathbb{X}_n)$), as being a conditional quantile, converges almost surely (respectively in probability) to the $(1-\alpha)$-quantile of the asymptotic null distribution $\mathcal{N}\left(0,\sigma^2_{P^1\otimes P^2}\right)$, denoted by $\Phi^{-1}_{0,\sigma^2_{P^1\otimes P^1}}(1-\alpha)$.

- As a direct consequence, combined with Slutsky's Lemma (see Appendix A.1.2, Proposition A.1.1), the probability that the bootstrap test wrongly rejects $(\mathcal{H}_0)$ satisfies

$$\mathbb{P}\big(\sqrt{n}U_{n,h}(\mathbb{X}_n) > q^*_{1-\alpha,n}(\mathbb{X}_n)\big) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}\left(Z_{P^1\otimes P^2} > \Phi^{-1}_{0,\sigma^2_{P^1\otimes P^1}}(1-\alpha)\right) = \alpha.$$

  The same holds for the permutation test, that is $\mathbb{P}\big(\sqrt{n}U_{n,h}(\mathbb{X}_n) > q^\star_{1-\alpha,n}(\mathbb{X}_n)\big) \underset{n\to+\infty}{\longrightarrow} \alpha$.

This being true for any underlying distribution P satisfying $(\mathcal{H}_0)$, the bootstrap (respectively permutation) test is thus asymptotically of prescribed size α.

On the other hand, the consistency of the tests furthermore requires the law of large numbers for U-statistics (recalled in Appendix A.3.2, Theorem A.3.1) which states that

$$U_{n,h}(\mathbb{X}_n) \underset{n\to+\infty}{\longrightarrow} \int_{\mathcal{X}^2\times\mathcal{X}^2} h(x,y)dP(x)dP(y), \ P\text{-a.s. in } (X_i)_i. \tag{0.5.5}$$

Moreover, since (0.5.2) (respectively (0.5.4)) stands as well under any alternative, the critical value $q^*_{1-\alpha,n}(\mathbb{X}_n)$ (respectively $q^\star_{1-\alpha,n}(\mathbb{X}_n)$) still converges to the Gaussian quantile, namely

$\Phi^{-1}_{0,\sigma^2_{P^1\otimes P^1}}(1-\alpha)$. Thus, under any reasonable alternative P such that the limit in (0.5.5) is positive,

$$\begin{aligned}\mathbb{P}\big(\sqrt{n}U_{n,h}(\mathbb{X}_n) > q^*_{1-\alpha,n}(\mathbb{X}_n)\big) &= \mathbb{P}\left(U_{n,h}(\mathbb{X}_n) > \frac{q^*_{1-\alpha,n}(\mathbb{X}_n)}{\sqrt{n}}\right)\\ &\underset{n\to+\infty}{\longrightarrow} \mathbb{P}\left(\int_{\mathcal{X}^2\times\mathcal{X}^2} h(x,y)dP(x)dP(y) > 0\right) = 1.\end{aligned}$$

The same holds for the permutation test replacing $q^*_{1-\alpha,n}(\mathbb{X}_n)$ by $q^\star_{1-\alpha,n}(\mathbb{X}_n)$. This being true for any underlying distribution P satisfying $\int_{\mathcal{X}^2\times\mathcal{X}^2} h(x,y)dP(x)dP(y) > 0$ the bootstrap (respectively permutation) test is thus consistent against any of these alternatives. Notice that the reconstruction ability is a much stronger property than just the consistency of the test which was already proved for instance by Hoeffding [84] for permutation tests.

The bootstrap and permutation tests with Monte Carlo approximation have also been proved to be asymptotically of prescribed size and consistent against the same alternatives. Intuitively, it is due to the convergence of the approximated quantiles to the limit of the exact ones, that is the quantile of the asymptotic null distribution.

0.5.2 Non-asymptotic study

These asymptotic results provide a first nice justification for the use of such approaches for testing independence. Yet, in many cases, and especially when applying such testing procedures to biological data as done in Chapter 2, the number of observation is often small due to economical or biological reasons, and purely asymptotic results may not be sufficient to theoretically validate the use of such procedures. This is why the non-asymptotic properties of the permutation tests have also been investigated. In this thesis, the non-asymptotic performance of a test is studied in terms of exact level and uniform separation rates, as presented in Section 0.4.2.

A well-known advantage of permutation approaches, is that, when correctly applied, they guarantee a non-asymptotic control of the first kind error rate by the prescribed level, as already stated above. Even though this result has been known for a very long time, for this thesis to be self-contained, let us prove, for instance, that the upper-tailed permutation test introduced in Section 0.4.3 is indeed exactly of prescribed level α. Consider P satisfying $(\mathcal{H}_0)$ and $\mathbb{X}_n$ a sample of i.i.d. random variables with distribution P. First recall that the exchangeability of the variables under $(\mathcal{H}_0)$ corresponds here to the fact that for all permutation π_n in $\mathfrak{S}_n$, $\mathbb{X}_n^{\pi_n}$ and $\mathbb{X}_n$ have exactly the same distribution. Thus

$$\mathbb{P}\big(\sqrt{n}U_{n,h}(\mathbb{X}_n) > q^\star_{1-\alpha,n}(\mathbb{X}_n)\big) = \frac{1}{n!}\sum_{\pi_n\in\mathfrak{S}_n}\mathbb{P}\big(\sqrt{n}U_{n,h}(\mathbb{X}_n^{\pi_n}) > q^\star_{1-\alpha,n}(\mathbb{X}_n^{\pi_n})\big).$$

Moreover, since the conditional quantiles are obtained from the order statistic when considering all permutations of $\mathbb{X}_n$, it is permutation invariant, that is, for all permutation π_n of $\{1,\dots,n\}$, $q^\star_{1-\alpha,n}(\mathbb{X}_n^{\pi_n}) = q^\star_{1-\alpha,n}(\mathbb{X}_n)$.

Hence, one obtains that

$$\begin{aligned}
\mathbb{P}\big(\sqrt{n}U_{n,h}(\mathbb{X}_n) > q^\star_{1-\alpha,n}(\mathbb{X}_n)\big) &= \frac{1}{n!}\sum_{\pi_n\in\mathfrak{S}_n}\mathbb{P}\big(\sqrt{n}U_{n,h}(\mathbb{X}_n^{\pi_n}) > q^\star_{1-\alpha,n}(\mathbb{X}_n)\big)\\
&= \sum_{\pi_n\in\mathfrak{S}_n}\mathbb{P}\big(\sqrt{n}U_{n,h}\big(\mathbb{X}_n^{\Pi_n}\big) > q^\star_{1-\alpha,n}(\mathbb{X}_n)\big|\Pi_n=\pi_n\big)\,\mathbb{P}(\Pi_n=\pi_n)\\
&= \mathbb{P}\big(\sqrt{n}U_{n,h}\big(\mathbb{X}_n^{\Pi_n}\big) > q^\star_{1-\alpha,n}(\mathbb{X}_n)\big)\\
&= \mathbb{E}\left[\mathbb{P}\big(\sqrt{n}U_{n,h}\big(\mathbb{X}_n^{\Pi_n}\big) > q^\star_{1-\alpha,n}(\mathbb{X}_n)\big|\mathbb{X}_n\big)\right]\\
&\leq \alpha,
\end{aligned}$$

by definition of the conditional quantile. This being true for all underlying distribution P satisfying the null hypothesis, the test is indeed exactly of level α.
In general, no equivalent result holds for the bootstrap approach. Therefore, when both approaches are available, the permutation one should always be preferred (see, e.g., [51]). This is the reason why we focused on the permutation test for the non-asymptotic study.

Moreover, it directly results from Romano and Wolf's lemma in [156, Lemma 1] that, thanks to the exchangeability of the random variables under the null hypothesis, the slightly modified permutation tests with Monte Carlo approximation, such as the one introduced in Section 0.4.3, are also exactly of prescribed level α, and this even with the Monte Carlo approximation step. The proof of this lemma is given by Blanchard et al. in [7].

Remains the study of the non-asymptotic control of the second kind error rate. The (asymptotic) minimax rates of testing independence between d real-valued random variables have been studied in the literature, initiated by Ingster [94], and then followed by Yodé [180] on Hölder classes with smoothness parameter ν with respect to different metrics. In particular, they are shown to be equal to

$$n^{\frac{-2\nu}{4\nu+d}} \text{ w.r.t. the } \mathbb{L}_2\text{-metric}, \quad \text{and} \quad \left(\frac{n}{\ln(n)}\right)^{\frac{-\nu}{2\nu+d}} \text{ w.r.t. the } \mathbb{L}_\infty\text{-metric}.$$

Then, Yodé [181] proposes an adaptive testing procedure, based on U-statistics which achieves the minimax rate w.r.t. the $\mathbb{L}_2$-metric up to a logarithm factor which is a reasonable and usual price to pay for adaptivity. Yet these approaches are purely asymptotic, and do not control the level for small sample size n (it is only shown to be controlled by a sequence that tends to α when n tends to infinity). However, these results are of main importances here since they provide the minimax rates of testing, which should be equivalent to the *minimax separation rate of testing*, defined in Section 0.4.2, which is our goal here.
Many other testing procedures that are adaptive in the minimax sense have been developed in the last decades in many other frameworks based on model selection procedures (see, e.g., [14, 55]) or thresholding methods (see, e.g., [167]). In this thesis, we adopt a thresholding method based on wavelet decomposition as detailed below.

Contribution of this thesis

As explained above, only the permutation test is studied from a non-asymptotic point of view. To bring back our study to more standard settings where the minimax rates are known, the analysis of the uniform separation rates in Chapter 4 is made in the density framework. In

this case, the null hypothesis is equivalent to the equality of the density function f with the product of its marginals $f^1 \otimes f^2$. For the purpose of adaptivity, we construct a multiple testing procedure based on aggregation of several single tests inspired by the wavelet thresholding methods investigated by Fromont et al. in [57] or Sansonnet and Tuleau-Malot in [160]. In wavelet settings, it is standard to consider regularity sets with smoothness properties such as Besov spaces (see, e.g., [13, 32, 55, 57, 60, 98, 99, 116, 160, 167]) or weak versions of them (see, e.g., [57, 58, 160]), and the $\mathbb{L}_2$-metric since all can be simply characterized thanks to the wavelet coefficients. More details can be found in Appendix A.4.
Fix two levels of errors α and β in $(0,1)$ and let $\mathbb{X}_n$ be a sample of random variables with distribution P.

The wavelet setting. Let $\{\varphi_\lambda\}_{\lambda\in\Lambda}$ be a wavelet basis of $\mathbb{L}_2([0,1]^2)$. In this thesis, we consider the Haar basis (see Chapter 4, Section 4.2.1, or Appendix A.4.1) for simplicity, but results may be generalized to more regular bases (see [39] or [126] for instance). For every index λ in Λ, introduce the corresponding coefficient $\beta_\lambda = \int_{[0,1]^2}[f(x) - f_1 \otimes f_2(x)]\varphi_\lambda(x)dx$ of the difference between the density function f and the product of its marginals $f_1 \otimes f_2$. Notice that under the null hypothesis, every coefficient β_λ is equal to zero.

The single tests. For each element of the basis, we construct a corresponding single coefficient two-tailed test based on the same permutation approach as in Section 0.4.3. More precisely, fix an index λ in Λ and define the new test statistic by $|T_\lambda(\mathbb{X}_n)|$ where $T_\lambda(\mathbb{X}_n) = U_{n,h_{\varphi_\lambda}}(\mathbb{X}_n)$ (see (0.4.2) and (0.4.3)). Since we aim at non-asymptotic results, the normalizing term $\sqrt{n}$ is dropped. In particular, $T_\lambda(\mathbb{X}_n)$ is an unbiased estimator of the coefficient β_λ. Then, the *single permutation two-tailed independence test* associated to λ rejects independence when

$$|T_\lambda(\mathbb{X}_n)| > q_{\lambda,1-\alpha}(\mathbb{X}_n),$$

where $q_{\lambda,1-\alpha}(\mathbb{X}_n)$ is the $(1-\alpha)$-quantile of the conditional distribution of the permuted test statistic $\left|T_\lambda(X_n^{\Pi_n})\right|$ given $\mathbb{X}_n$, where Π_n is a uniform permutation of $\{1,\dots,n\}$ independent of $\mathbb{X}_n$.

As expected, by construction, the test is exactly of prescribed level α. Recall that the aim is to upper bound the uniform separation rate, up to a constant, by the *minimax separation rate of testing* over regularity spaces with respect to the $\mathbb{L}_2$-metric. As a first step, we exhibit a condition on the alternative f which guarantees a control of the second kind error rate by the prescribed value β. More precisely, since the coefficient $\beta_\lambda = \beta_\lambda(f)$ is equal to zero under $(\mathcal{H}_0)$, the bigger it is, the further the alternative f is from $(\mathcal{H}_0)$ and the better the test is able to detect it. Hence, we find a threshold s, depending on n, α, β, φ_λ and $\|f\|_\infty$ such that if $|\beta_\lambda| \geq s$, then the probability under the alternative f that the test wrongly accepts $(\mathcal{H}_0)$ is controlled by β. This result is based on a concentration inequality for randomly permuted sums presented in Section 0.5.3.

The aggregated test. In order to avoid the tricky question of the choice of the coefficient λ, and to be able to detect more general forms of dependences, the idea is to aggregate the single coefficient tests up to some scale $\tilde{J}$ in the wavelet decomposition. As usually in multiple testing theory, aggregating tests requires a correction of the single levels. Here, we consider the same correction as for instance in [14, 57, 58] that is the sharpest correction of the individual levels allowing to apply simultaneously all single coefficient tests with global level α (see Chapter 4,

Section 4.2.3). Then, the *aggregated permutation two-tailed independence test* Δ_α rejects the null hypothesis if at least one of the single permutation two-tailed independence tests (with corrected level) rejects it.

By construction of the single level corrections, the aggregated test is exactly of level α. Then, for the study of the uniform separation rates, we consider a certain regularity set $\mathcal{BW}_{\delta,\gamma,\infty}(R, R', R'')$ properly defined in Chapter 4, and informally being the intersection of a Besov body $\mathcal{B}^\delta_{2,\infty}(R)$ with smoothness parameter $\delta > 0$ and radius $R > 0$, a weak Besov body $\mathcal{W}_\gamma(R')$ with smoothness parameter $\gamma > 0$ and radius $R' > 0$ (see Appendix A.4.2), and the $\mathbb{L}_\infty$-ball with radius $R'' > 0$. Finally, we obtained that if the maximal scale $\tilde{J}$ is equal to $\lfloor \log_2 \sqrt{n/\ln(n)} \rfloor$, and $\delta \geq \gamma/(\gamma+1)$, then, for n large enough,

$$\rho\left(\Delta_\alpha, \mathcal{BW}_{\delta,\gamma,\infty}(R, R', R''), \beta\right) \leq C(\alpha, \beta, \delta, \gamma, R, R', R'') \left(\frac{n}{\ln(n)}\right)^{\frac{-\gamma}{2\gamma+2}},$$

for some positive constant $C(\alpha, \beta, \delta, \gamma, R, R', R'')$. Even though the minimax separation rate over such two-dimensional regularity sets with respect to the $\mathbb{L}_2$-metric have not been derived yet, this uniform separation rate seems to be optimal in the minimax sense, in view of the literature and more precisely the results of Ingster [94] and Yodé [180, 181]. Moreover, since the testing procedure does not depend on the knowledge of the smoothness parameters δ and γ, it seems to be adaptive in the minimax sense over such regularity spaces. A detailed discussion is done in the introduction of Chapter 4.

0.5.3 Concentration inequalities: an efficient tool

Concentration inequalities are central and powerful tools for the study of uniform separation rates, as, for instance, in [58, 160]. They allow to control the probability that a random variable $Z_n = \zeta(Y_1, \ldots, Y_n)$, that is a function of one or several random variables $Y_1, \ldots, Y_n$, moves away from its median or its mean, by a prescribed real number. In particular, they lead to precise upper-bounds for the permutation conditional quantiles.

The literature on concentration inequalities is colossal (see the books of Ledoux [114], Massart [123], or the very recent one of Boucheron, Lugosi, and Massart [26] for some great reviews). They initially appeared for sums of independent random variables, such as, for instance, Hoeffding's [86], Bennett's [18] and Bernstein's [19] inequalities. Of main interest here, let us recall the famous Bernstein inequality stated for instance by Massart in [123, Proposition 2.9 and Corollary 2.10].

Theorem 0.5.1 (Bernstein's inequality)**.** *Let $X_1, \ldots, X_n$ be independent real-valued random variables. Assume that there exists some positive numbers v and c such that*

$$\sum_{i=1}^n \mathbb{E}\left[X_i^2\right] \leq v \quad \text{and} \quad \forall k \geq 3, \ \sum_{i=1}^n \mathbb{E}\left[(X_i)_+^k\right] \leq \frac{k!}{2} v c^{k-2},$$

where $(\cdot)_+ = \max\{\cdot, 0\}$ denotes the positive part.

Let $S = \sum_{i=1}^n (X_i - \mathbb{E}[X_i])$, then for every positive x,

$$\mathbb{P}\left(S \geq \sqrt{2vx} + cx\right) \leq \exp(-x). \tag{0.5.6}$$

Moreover, for any positive t,

$$\mathbb{P}(S \geq t) \leq \exp\left(-\frac{t^2}{2(v+ct)}\right). \tag{0.5.7}$$

Notice that both forms of Bernstein's inequality appear in the literature. Yet, due to its form, (0.5.6) is rather preferred in statistics for the control of the quantiles, even though (0.5.7) is more standard.

A random variable $Z_n = \zeta(Y_1 \ldots, Y_n)$ is said to satisfy a *Bernstein-type inequality* if there exists some positive numbers c and v respectively depending on the distribution of the Y_i's and precisely the $\mathbb{E}\left[Y_i^2\right]$'s, and some universal constants c_0 and c_1 such that, for all positive x,

$$\mathbb{P}\left(Z_n \geq \sqrt{2vx} + cx\right) \leq c_0 \exp\left(-c_1 x\right),$$

or more classically, if for all positive t,

$$\mathbb{P}(Z_n \geq t) \leq c_0 \exp\left(-\frac{c_1 t^2}{2(v+ct)}\right).$$

Furthermore, in the present work, we are interested in studying the concentration of permuted sums, which are sums of dependent variables (namely $Y_i = a_{i,\Pi_n(i)}$, where $a_{i,j}$ are real-valued numbers and Π_n is a uniform permutation of $\{1, \ldots, n\}$). Concentration inequalities for functions of dependent random variables have been developed from different approaches.

A first approach, as explained in [168], is to decompose Z_n as a sum of a martingale difference sequence and apply martingale inequalities. For instance, exponential inequalities for sums of weakly dependent variables have been obtained by Delyon in [42], based on the martingale theory.

Another approach is based on Stein's method which consists in approximating complicated distributions (such as the distribution of a sum of dependent variables) by a much simpler and tractable one through the identification of a characterizing operator. Based on such approach, Chatterjee [33, Proposition 1.1] obtains a concentration inequality for permuted sums recalled in Chapter 3, Theorem 3.1.3 stating that for all positive t,

$$\mathbb{P}(|Z_n - \mathbb{E}[Z_n]| \geq t) \leq 2\exp\left(-\frac{t^2}{4\mathbb{E}[Z_n] + 2t}\right).$$

Instead of investigating such approaches, we specially focus on concentration inequalities for random permutations, and in particular the fundamental inequality of Talagrand [168, Theorem 5.1] which states the following, as recalled in Chapter 3, Theorem 3.1.1.

First, Talagrand introduces a notion of distance between a permutation π_n in $\mathfrak{S}_n$ and a subset A of $\mathfrak{S}_n$. To do so, he reduces the set of interest, namely $\mathfrak{S}_n$, to a simpler one, that is $[0,1]^n$, by considering

$$U_A(\pi_n) = \left\{s \in \{0,1\}^n \ ; \ \exists \tau \in A \text{ satisfying } \forall 1 \leq i \leq n, \ s_i = 0 \implies \tau(i) = \pi_n(i)\right\}.$$

In particular, the permutation π_n belongs to A if and only if the null vector belongs to $U_A(\pi_n)$. He thus defines the distance between π_n and A as the square of the standard ℓ_2-distance between the null vector 0 and the convex hull of $U_A(\pi_n)$ in $[0,1]^n$, denoted $V_A(\pi_n)$, that is

$$f(A, \pi_n) = \min\left\{\sum_{i=1}^n v_i^2 \ ; \ v = (v_i)_{1 \leq i \leq n} \in V_A(\pi_n)\right\}.$$

In particular, it is important to notice that the permutation π_n belongs to A if and only if $f(A, \pi_n) = 0$.
Then, Talagrand's inequality for random permutations states that, if P_n denotes the uniform distribution on $\mathfrak{S}_n$, then

$$\int_{\mathfrak{S}_n} \exp\left(\frac{1}{16} f(A, \pi_n)\right) dP_n(\pi_n) \leq \frac{1}{P_n(A)}.$$

Concentration inequalities are then derived from Markov's inequality, and in particular, for all $t > 0$,

$$P_n\left(\pi_n \; ; \; f(A, \pi_n) \geq t^2\right) \leq \frac{\exp\left(-t^2/16\right)}{P_n(A)}.$$

Based on this inequality, many exponential concentration inequalities have been obtained for instance by McDiarmid [124] under Lipschitz-type conditions, or more recently Adamczak et al. [2] under convex Lipschitz conditions recalled in Chapter 3, Theorem 3.1.2.
Moreover, Talagrand's inequality for random permutations usually leads to concentration inequalities around the median, whereas we are more interested in concentration around the mean (since in the *Linear case*, the permuted test statistic is a centered permuted sum, as shown in (0.5.3)). However, Ledoux [114] shows that mean and median are equivalent under exponential tails. Hence, the inequalities around the mean are the same, up to constants, as the ones around the median.

Contribution of this thesis

Let $\{a_{i,j}\}_{1 \leq i,j \leq n}$ be a collection of real numbers and denote for each permutation π_n in $\mathfrak{S}_n$, the corresponding sum $Z_n(\pi_n) = \sum_{i=1}^n a_{i,\pi_n(i)}$. We aim at finding an exponential concentration inequality for the randomly permuted sum $Z_n(\Pi_n)$. McDiarmid's [124] or Adamczak et al.'s [2] results seem inadequate here since their Lipschitz-type conditions may be very restrictive. More particularly, in our study of uniform separation rates in Chapter 4, these conditions are not satisfied. Hence, as explained in Chapter 3, the idea is thus to exploit the attractive summation form. Based on Talagrand's inequality for random permutations, we sharpen Chatterjee's [33] result by recovering a variance term instead of an expectation term in the exponential.
We follow Adamczak et al.'s [2] idea and apply Talagrand's inequality for random permutations. In particular, we first obtain a rough inequality for $\sqrt{Z_n(\Pi_n)}$ around its median in the non-negative case ($a_{i,j} \geq 0$ for all $1 \leq i, j \leq n$) which is insufficient. The idea is to plug this first inequality back into the proof in order to sharpen the result. The transition from the median to the expectation results from Ledoux's [114] trick. It unfortunately leads to drastically large constants. At the end, we obtain a *Bernstein-type inequality* in the general (not necessarily non-negative) case, that is for all positive x,

$$\mathbb{P}\left(|Z_n - \mathbb{E}[Z_n]| \geq 2\sqrt{2\left(\frac{1}{n}\sum_{i,j=1}^n a_{i,j}^2\right)x} + 2\max_{1 \leq i,j \leq n} |a_{i,j}|\, x\right) \leq c_0 \exp\left(-c_1 x\right),$$

where c_0 and c_1 are universal positive constants.

0.6 Application to synchrony detection in neuroscience

0.6.1 The neuroscience context and motivation

Understanding how the brain works is a great challenge. Neural information is transmitted through the brain as electric and chemical signals thanks to neurons which are excitable cells. In broad outline, neurons communicate as follows. The dendrites receive synaptic input signals from one or various "previous" neurons. These input signals increase the membrane electric potential of the neuron at the level of the cell body. If the membrane potential exceeds some excitation threshold, it rapidly increases and falls. This brief and stereotyped depolarization, called an *action potential*, propagates along the axon, and the neuron is thus said to *fire*. When the action potential reaches the axon terminals, the chemical synapses send new signals to the dendrites of one or several postsynaptic neurons.

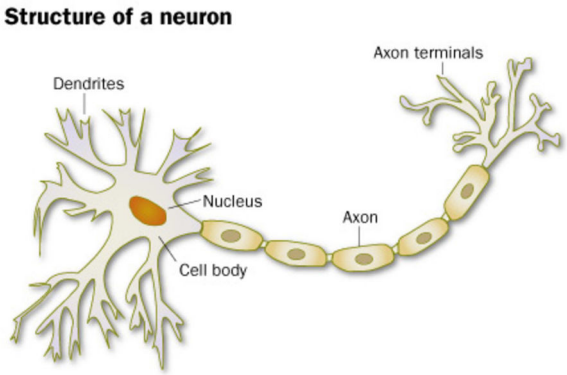


Figure 2 – Scheme of a neuron.

It is now accepted that action potentials are one of the main components of the brain activity (see [166]). The set of time occurrences of the action potentials of a neuron is called a *spike train* (the *spikes* being the time occurrences themselves). It is commonly modeled by point processes (see Section 0.6.2), each point representing a spiking time. The expected number of spikes per unit of time is called firing rate. Action potential are triggered mainly in two different ways; either the "previous" neurons increase (independently) their instantaneous firing rate, which increases the frequency of the input signals (see [15]), or they coordinate their activity, such that the input signals nearly appear at the same time (see [79]). The second phenomenon, called *synchronization*, needs less input signals, uses less energy, and increases the speed of transmission of the neural information. It is now commonly admitted to play an important role in the neural activity (see [171]).

A first step in understanding the neuronal code is to be able to detect such synchronization between two or more neurons. On the one hand, this requires the development of experimental methods able to record simultaneously several neurons. It is nowadays possible thanks to techniques such as the combination of multi-microelectrodes which record extracellular neural activities, and spike sorting algorithms allowing to isolate the signals of each concerned neurons. On the other hand, various statistical methods for synchrony analysis have been developed in the neuroscience literature (see, for instance, the reviews of Harrison et al. [77] or Grün [68]). They are usually based on the notion of *coincidence* between two or more neu-

rons, which occurs when the concerned neurons fire nearly at the same time. Since during a synchronization, the neurons synchronize their activities, their spike trains become dependent. The general idea is thus to compare the observed number of coincidences to what is expected under the independence hypothesis; if it is larger than expected, the correlation is said to be *excitatory*, and in the opposite, if it is smaller, the correlation is said to be *inhibitory*.

A popular tool for illustrating such synchrony phenomenon between two neurons is the *cross-correlogram* (see [133] or [77] for a detailed description of this method). In a few words, it represents the histogram of the lags between spiking times of two different neurons. A narrow peak in the cross-correlogram represents a correlation between the two corresponding neurons since it shows that the neurons tend to fire at some privileged time lapse. A synchronization is usually identified by such a peak for a lag close to zero. Yet, these methods may not be precise enough, since they are mostly visual and may thus be seen as descriptive statistics. In particular, alone, they do not provide the possibility of evaluating the statistical significance of the obtained results. Harrison et al. develop in [77], a bootstrap-based method which, under a parametric Poisson model, provides confidence regions for cross-correlograms (one rejects independence if the cross-correlogram exits the region). Yet the Poisson assumption is questionable and sometimes fails to fit real data (as for instance in [10, 53, 147]). Many other methods have been introduced.

One of the most well-known method to detect synchronizations is the famous Unitary Events (UE) analysis method, introduced by Grün and her collaborators in the late 1990's (see [67, 71]). This method allows to detect functional groups or patterns of two or more neurons that are correlated (in an excitatory or inhibitory manner). More precisely, when observing L different neurons, one may want to detect particular forms of patterns (for instance, all neurons fire together corresponds to the pattern $\{1, \dots, L\}$, which is our case with $L = 2$, or neurons 1, 3 and 4 fire together, corresponding to $\{1, 3, 4\}$). There are 2^L possible patterns, and each pattern is studied separately. Now fix a pattern. One of the main advantages of the original UE method is that, given n observations of L spike trains recorded simultaneously, it quantifies the degree of dependence assessing p-values as follows.

Description of the original UE method.

- First, a pre-processing *binning* and *clipping* of each spike train is done: chose an arbitrary bin length δ, corresponding to a time discretization (binning), and reduce the spike count within a bin to a binary event (clipping): assign for all bin, "zero" if no spike occurred during the bin, and "one" if at least one spike occurred. The spike train can thus be seen as a finite dimensional random vector with coordinates in $\{0, 1\}$, and dimension equal to the number of bins.

- For each trial, a bin is said to contain a *coincidence* between the neurons of the fixed pattern at that trial if all the corresponding spike trains have a "one" in that bin. Then, the observed total number of coincidences n_{emp} corresponds to the sum over the trials of the numbers of bins containing a coincidences between the neurons of the pattern. A more precise definition in the case of $L = 2$ neurons is given in Chapter 2, Definition 2.2.1.

- As already mentioned, the general idea is to compare the observed total number of coincidences to the one expected under independence of the neuronal activities. To do so, the expected total number of coincidences is approximated by a function of the joint probability that all neurons in the pattern have a "one" in a bin. Under independence, this

joint distribution is just the product of the marginals, hence estimating those marginals leads to a good estimate of the expected total number of coincidences n_{exp}.

- Assuming moreover that the total number of coincidences follows a Poisson distribution, the assessed p-value is equal to the probability that a Poisson random variable with parameter n_{exp} is greater than the observed one n_{emp}.

  Then the test rejects independence if the p-value is either smaller than the prescribed level α (inferring excess synchrony, and thus an excitatory correlation) or larger than $1-\alpha$ (inferring significant missing coincidences, and thus an inhibitory correlation).

- Finally, to detect when the neurons of the pattern coordinate their activity, the test is simultaneously performed on many different windows, and the detected synchronizations correspond to windows on which independence is rejected.

Several improvements of the original method have been investigated and are summarized in the introduction of Chapter 2.
Firstly, the binning of the data may involve drastic information loss in synchrony detection (see [72]), since the coincidence count highly depends on the length and the origin of each bin. Notably, coincidences such that the spike from the first neuron is close to the end of a bin, whereas the spike from the second neuron is close to the beginning of the next bin, cannot be detected, however close they are. To overcome such information loss, a new coincidence count without coarse binning, namely the *multiple shift coincidence count*, is introduced by Grün et al. in [72] for discretized data, and has been generalized to the *delayed coincidence count*, denoted φ_δ^{coinc}, for two point processes by Tuleau-Malot et al. in [170] (see Chapter 2, Definition 2.2.2 for more details) and then, to three or more point processes by Chevallier and Laloë in [35].

Secondly, the Poisson assumption concerning the coincidence count does not seem realistic.

Thirdly, in multiple testing, the multiplicity of the tests should be taken into account. For instance, if m independent single tests are performed simultaneously at individual size α, then, globally, the probability of wrongly rejecting at least one true null hypothesis is equal to $(1-(1-\alpha)^m)$, which is equivalent to $m\alpha$ when α tends to zero, and is thus much larger than α if m is large. Hence each individual levels need to be corrected. Many methods with different corrections have been proposed in the literature. Among them lies Benjamini and Hochberg's [16] famous procedure detailed in Section 0.6.4.

Tuleau-Malot et al. [170] propose a theoretically justified method for two point processes (generalized in [35] to three or more neurons) which does not suffer from the previous drawbacks. This method, called *Multiple Tests based on a Gaussian Approximation of the Unitary Events* (MTGAUE), is based on a Gaussian approximation of the centered normalized delayed coincidence count. More precisely, assume that the observed trials are n i.i.d. copies $(X_1,\dots,X_n)$ of a couple of point processes $X=(X^1,X^2)$.

Description of the MTGAUE method.

- Fix a window $W=[a,b]$ and compute the empirical mean of the number of (delayed) coincidences φ_δ^{coinc} on W, that is,

$$\bar{Z}_n = n^{-1}\sum_{i=1}^{n}\varphi_\delta^{coinc}(X_i^1,X_i^2).$$

- If X^1 and X^2 are two independent homogeneous Poisson processes with intensities λ^1 and λ^2 (see Section 0.6.2), then, the expected number of coincidences between X^1 and X^2 can be expressed in terms of λ^1 and λ^2, and is denote here by

$$\mathbb{E}\left[\varphi_\delta^{coinc}(X^1, X^2)\right] = m(\lambda^1, \lambda^2).$$

  The plug-in step consists in replacing the unknown parameters λ^1 and λ^2 by their unbiased estimators $\hat{\lambda}^1$ and $\hat{\lambda}^2$ defined by

$$\hat{\lambda}^j = \frac{1}{n(b-a)} \sum_{i=1}^{n} N_{X_i^1}([a, b]),$$

  where $N_{X_i^1}([a, b])$ denotes the number of points of X_i^1 in $[a, b]$.

- The test statistic T_n, which is equal to $(\bar{Z}_n - m(\hat{\lambda}^1, \hat{\lambda}^1))$ correctly normalized, is shown to converge in distribution to the standard Gaussian distribution $\mathcal{N}(0, 1)$ under the null hypothesis. Hence the symmetric test rejects independence on the window W if $|T_n|$ is greater than the $(1 - \alpha/2)$-quantile of $\mathcal{N}(0, 1)$.

- Finally, as for the original UE method, to detect when the neurons coordinate their activity, the test is simultaneously performed on many different windows. Yet, in order to take into account the multiplicity of the tests, a Benjamini and Hochberg procedure is applied (see Section 0.6.4).

However, this method, as the original UE method is still based on very restrictive assumptions on the spike trains, such as a strong stationarity assumption, and may therefore fail on experimental data.

To depart from these stationarity assumption, several procedures have also been introduced in the literature, based on resampling approaches. In particular, two different bootstrap type procedures have been performed.
On the one hand, non-parametric models, such as inhomogeneous Poisson processes, have been used. Their intensity is non-parametrically estimated, and processes with this estimated intensity are simulated to reconstruct the null distribution. This is a particular case of parametric bootstrap. This approach is used, for instance, by [175, 176, 103], possibly using more involved models, with conditional intensities given the history of the process, such as inhomogeneous Markov interval models, Gamma process models or loglinear models.
On the other hand, neuroscientists have developed other resampling-based methods to be able to reconstruct the null distribution without making model assumptions. These methods include the whole family of surrogate data methods (see [120] for a review), such as the spike time randomization, the spike train dithering or the spike time dithering methods (see [121]) which are surrogate data generating methods destroying possibly existing coincidences by breaking the internal structure of spike trains across time. However, we do not know any mathematical structures that would theoretically legitimate those surrogate across time methods, except maybe renewal processes.
One could also consider the bootstrap-based method rapidly introduced by Ventura in [175] and based on *joint peri-stimulus time histograms* (jPSTH) (see [3]). More precisely, consider the binning processing introduced as a first point in the description of the Unitary Events method, with bin length $\varepsilon > 0$ small enough so that no bin can contain more than one spike.

Then, the *peri-stimulus time histograms* (PSTH) of a neuron is the bin by bin sum across trials of its binned spike trains divided by the number of trials. It is a natural estimator of the instantaneous firing rate of a neuron, that is the (theoretical) number of spikes per unit of time. The jPSTH corresponding to two neurons is the PSTH of the bin by bin product of the individual discretized spike trains. To solve the non-stationarity effect, her test statistic, based on the jPSTH as described above and the products of the marginal PSTHs, depends on time t. Notice that it can be written as a particular case of our test statistic $U_{n,h_{\varphi_t}}$ defined by both (0.4.2) and (0.4.3), where $\varphi_t(X^1, X^2) = \mathbb{1}_{N^1_X([t,t+\varepsilon[)=1,\ N^2_X([t,t+\varepsilon[)=1}$ for each t corresponding to a bin. Its distribution under the independence hypothesis is obtained from a non-parametric bootstrap method (with or without a Monte Carlo approximation) which consists in resampling independently from the trials of each neuron, as introduced in Section 0.4.3. To solve the multiplicity issue due to the time dependency, independence is rejected when the test statistic, as a curve, exits the joint null estimated envelope. Yet, in view of Chapter 1, and to fully justify this approach from a theoretical point of view, one needs to prove that the bootstrap approach is not only able to reconstruct the null distribution of each $U_{n,h_{\varphi_t}}$ for each bin t, but of $\sup_t \left|U_{n,h_{\varphi_t}}\right|$. Moreover, the choice of the bin length ε is very tricky. Indeed, choosing ε too small might lead to null $\varphi_t(X^1, X^2)$ for most of the bins, whereas choosing ε too large might suffer, as any binning method, from a loss in synchrony detection.

Another popular surrogate method is the *trial-shuffling* method due to Pipa et al. [137, 138], which destroys coincidences by breaking the couples of spike trains across trials. Intuitively, the trial shuffling approach consists in sampling with replacement in all couples of spike trains obtained from different trials, that is with the notation introduced above, $\{(X^1_i, X^2_j)\}_{i\neq j}$ (see Chapter 2, Section 2.3.2 for more details). Even though it is usually considered in the neuroscientific literature as a permutation method, it does not correspond to the permutation approach introduced in this thesis (see Section 0.4.3). Indeed, on the one hand, the permuted sample may contain two spike trains from the same trial (as soon as the random permutation has a fixed point), which is forbidden in trial-shuffling due to the resampling method, and on the second hand, it consists in resampling *with* replacement, which is closer to general bootstrap methods than permutation ones. More precisely, this method is a non-parametric bootstrap method (as the one in [175], and unlike those in [175, 176, 103]), and therefore, does not need any model assumption. However, it is based on binned data and might therefore suffer from a loss of synchrony detection, as, for instance, the original UE method. Moreover, in view of the analysis of the centering issue done in Chapter 2, bootstrap methods applied to non centered statistics do not lead to correct approximations of the null distribution. In particular, the *trial-shuffling* method directly bootstraps the number of coincidences, and therefore does not reconstruct the desired null distribution (see Chapter 2).

0.6.2 Modeling by point processes

Point processes appear to be a natural way of modeling neuronal activity (see, for instance, the nice introduction to point processes in Kass et al.'s book [102, Chapter 19]). Indeed, even though the real recorded spike trains are discretized in time due to the record resolution, and thus belong to finite dimensional spaces, the dimension of these spaces is so huge (from ten thousands to a million) that it is neither realistic nor reasonable to model them by finite dimensional vectors, and point processes seem more adapted.

A quick introduction to point processes

Point processes have been largely studied during the last decades and applied in many different frameworks such as economics, epidemiology, seismology, telecommunications, computational neuroscience and many others. An introduction to the theory of point processes can be found in the book of Daley and Vere-Jones [38].
A point process X on a measurable set I is a random countable subset of I. Denote by $\mathcal{X}$ the set of all their possible values. For biological reasons, we only focus on almost surely finite point processes. They are characterized by the fact that for all bounded subset J of I, the number of points of the process X in J, denoted by $N_X(J)$ is almost surely finite. In particular, such point processes cannot have any accumulation point.
For all x in $\mathcal{X}$, define the associated counting measure by $dN_x = \sum_{u\in x} \delta_u$, satisfying for all real-valued, measurable function f on I,

$$\int_I f(s) dN_x(s) = \sum_{u\in x} f(u).$$

In the following, since we are interested in point processes indexed by time, we focus on the particular case where I is of the form $[0,T]$ with $0 < T < +\infty$, or equal to $\mathbb{R}_+$. In this case, the point process X can be identified with its counting process, defined by

$$N_X : \begin{pmatrix} I & \longrightarrow & \mathbb{R} \\ t & \longmapsto & \int_0^T \mathbb{1}_{s\leq t}\, dN_X(s) \end{pmatrix}.$$

In particular, $N_X(t)$ is a random variable that counts the number of points of X in $[0,t]$. The process $(N_X(t))_{t\geq 0}$ is non-negative, non-decreasing, piecewise constant, each jump being equal to one and occurring at the points of the process. It is moreover càdlàg (for "continue à droite, limitée à gauche" in french), that means it is right-continuous and limited on the left. Through the identification between a point process X and its associated counting process N_X, the set of point processes $\mathcal{X}$ can be endowed with a metric $d_{\mathcal{X}}$ issued from the Skorohod topology thanks to its embedding in the set of càdlàg functions (see [22] or Appendix A.2.2).

The distribution of a point process may be defined in several ways. First, it may be represented by its probability measure on Borel sets of $(\mathcal{X}, d_{\mathcal{X}})$, as done in this thesis (see Section 0.4.2 in the point process framework). Yet, it seems difficult to characterize simple and usual point processes thanks to this approach.
Another, and much more typical tool for characterizing the distribution of a point process is the conditional intensity $\lambda(t)$ given the past. Basically, it depends on the history of the process until time t, coded in its natural filtration, that is the σ-algebra $\mathcal{F}_{t^-}$ generated by $\{N_X(s), s < t\}$. Intuitively, the conditional intensity λ is a non-negative measurable function such that $\lambda(t)dt$ is the conditional probability of finding a new point in $[t, t+dt)$ given $\mathcal{F}_{t^-}$, that is

$$\lambda(t)dt \approx \mathbb{E}\left[dN_X(t)|\mathcal{F}_{t^-}\right].$$

It is not in the scope of this thesis to give precise definitions of such tools, and the interested reader could refer to the books of Brémaud [28] or Daley and Vere-Jones [38]. A great advantage of this definition is that, given the conditional intensity, it is possible to construct a point process with such intensity in a simple way thanks to Ogata's [129] thinning procedure, described in Appendix A.2.1. In neuroscience, this notion corresponds to the instantaneous firing rate, which also looks at the infinitesimal probability that the neuron triggers a spike.

Some typical examples

Homogeneous Poisson processes. The simplest example of point processes is the well-known *homogeneous Poisson process*, and is defined as follows in [157, Definition 5.1].

Definition 0.6.1. A homogeneous Poisson process X with rate $\lambda > 0$ on I is defined by

- $N_X(0) = 0$,
- X has independent increments, that is, for all non-negative integer l, and all disjoint measurable subsets $A_1, \dots, A_l$ of I, $N_X(A_1), \dots, N_X(A_l)$ are independent variables,
- for all $s > 0$, the number of points in any interval with length s is Poisson distributed with parameter λs, that is, for all $t > 0$ such that $[t, t+s)$ is included in I, for all non-negative integer k,

$$\mathbb{P}(N_X([t, t+s)) = k) = \frac{(\lambda s)^k}{k!} \exp(-\lambda s).$$

The Poisson process has many advantageous properties. Among them, it has stationary increments, that is the distribution of the number of points occurring in any interval only depends on the length of this interval, and not its position in I. Moreover, the intensity of a homogeneous Poisson process is equal to its rate, and is thus a deterministic constant function. This notably simplifies the study of such point processes and is the reason why they are vastly encountered in the neuroscience literature (see, for instance, [67, 170, 35]) even though they are too simple to be realistic models.

Inhomogeneous Poisson processes. A first generalization of homogeneous Poisson process consists in removing the stationarity property, which leads to *inhomogeneous Poisson processes*.

Definition 0.6.2. An inhomogeneous Poisson process X with mean measure μ containing no atoms on I is defined by

- X has independent increments (see Definition 0.6.1),
- for any measurable subset A of I, the number of points in A is Poisson distributed with parameter $\mu(A)$, that is for all non-negative integer k,

$$\mathbb{P}(N_X(A) = k) = \frac{\mu(A)^k}{k!} \exp(-\mu(A)).$$

Generally, the measure μ is assumed to be absolutely continuous with respect to the Lebesgue measure, and one can thus denote, $\mu(dt) = \lambda(t)dt$. The function λ is called the intensity of the process.

First notice that the intensity of a inhomogeneous Poisson process is equal to its conditional intensity. Moreover, if it is constant, one recovers a homogeneous Poisson process.

The main advantage of Poisson processes (homogeneous or not) are that their conditional intensity is deterministic, and does not depend on the history of the process. This property makes them mathematically appealing. In the literature, they are often used to model for instance radioactive particles emission times, calls in a telephone exchange, or the arrival times of clients at some stand or ticket office. Yet, such models are not realistic in neuroscience as they fail to model biological features such as the refractory periods, which are short delays after each spike in which the neuron cannot fire again due to biological reasons.

Hawkes processes and Poisson interaction models. A generalization of Poisson processes is the *Hawkes process* which enables to model self-excitatory features, or on the contrary self-inhibiting ones such as the refractory period in neuroscience. Introduced by Hawkes [78] to model replicas of earthquakes in seismology (see [130, 177]), Hawkes processes have been applied in diverse areas such as genomics (see [73, 149]), finance (see [11, 12]) and recently neuroscience (see [113, 134, 147, 148, 160]).

Definition 0.6.3. A univariate Hawkes process is a point process X with conditional intensity of the form

$$\lambda(t) = \Phi\left(\int_0^t h(t-s)dN_X(s)\right) = \Phi\left(\sum_{u\in X, u<t} h(t-u)\right), \tag{0.6.1}$$

where Φ is a non-negative function, and $h : \mathbb{R}_+ \to \mathbb{R}$ is called the self-interaction function. Throughout this thesis, Φ is taken as $\Phi(\cdot) = \max\{0, \mu + \cdot\}$, where $\mu > 0$ is called the spontaneous contribution.

The dependency in the past is coded in the integral/sum term in (0.6.1). Each point of the process u occurring before t gives an "opinion" on the appearance of a new point at time t through the function h depending on its age $t-u$. In particular, if $h(t-u)$ is positive, then the "opinion of u" increases the instantaneous probability of finding a new point at time t and has thus an excitatory effect, whereas if it is negative, then u has an inhibitory effect. In particular, if h is very negative on a small neighborhood of zero, say $[0, r]$, such that for any point u of the process, $\mu + \int_0^t h(t-s)dN_X(s)$ is negative for all t in $[u, u+r]$, then the intensity is equal to zero on each $[u, u+r]$ and hence, no point can occur at distance less than r from an existing one. This exactly models the refractory period. Finally, all points u of the process such that $h(t-u) = 0$ have "no opinion" on the appearance of a new point at time t. And in particular, if h is equal to the null function, then one recovers a homogeneous Poisson process.
Since interactions between neurons are at the root of synchrony analysis, dependent multivariate Hawkes processes have also been considered in the neuroscience literature.

Definition 0.6.4. A bivariate Hawkes process is a couple of (interacting) Hawkes processes (X^1, X^2) where for $i = 1, 2$, the conditional intensities λ^i of X^i is of the form

$$\lambda^i(t) = \Phi_i\left(\sum_{j=1}^{2}\int_0^t h_{j\to i}(t-s)dN_{X^j}(s)\right),$$

with Φ_i are non-negative functions (taken as in Definition 0.6.3), and $h_{j\to i} : \mathbb{R}_+ \to \mathbb{R}$ is called self-interaction function if $j = i$ and interaction function otherwise.

In addition to its dependence with respect to its own history, each process of the bivariate Hawkes process, say X^1, depends on the history of the other process, say X^2, through the interaction function, namely $h_{2\to 1}$. In particular, in neuroscience, if $h_{2\to 1}$ is positive, then neuron 2 excites the activity of neuron 1, whereas if it is negative, then neuron 2 inhibits the activity of neuron 1. This definition can be generalized to more than 2 processes, and is then referred to as multivariate Hawkes processes.
The *Poissonian interaction model* introduced by Sansonnet and Tuleau-Malot in [160] is another model for spike trains encountered in synchrony analysis. It is a particular case of bivariate Hawkes processes (X^1, X^2), where both self interaction functions $h_{1\to 1}$ and $h_{2\to 2}$,

and the interaction function $h_{2\to1}$ are equal to the null function. In this case, the "parent" process X^1 is simply a homogeneous Poisson process, and the "child" process X^2 only depends on the history of the "parent" process. Testing independence reduces to testing the nullity of the function $h_{1\to2}$.
Moreover, a Hawkes process for which the function Φ in Definition 0.6.4 is the exponential function can be seen as a particular case of the *generalized linear models* (see, for instance, [136, 169]).

However, as already mentioned, neither of these models, nor any other, is commonly accepted in the neuroscience literature. Nevertheless, we use those simple models in the simulation studies.

0.6.3 A quick introduction to Multiple Testing

To treat the synchrony detection issue, the usual approach consists in testing independence simultaneously on many sliding windows. To do this, multiple testing procedures are thus necessary, and a brief introduction to such methods is given here. The multiple testing problem arises when one wants to test simultaneously m null hypothesis, say $(\mathcal{H}_{0,1}), \dots, (\mathcal{H}_{0,m})$ with some global error rate controlled by some level α.
A multiple testing procedure is a function of the observation which returns a subset of $\{1, \dots, m\}$ corresponding to the indices of the rejected null hypotheses. In the following, a *positive* refers to a null hypothesis that is declared significant and is thus rejected, and on the contrary, a *negative* refers to a null hypothesis which is not significant in view of the data and is therefore accepted. Define then a *false positive* (respectively a *false negative*) a null hypothesis that is wrongly rejected (respectively accepted). Hence, in analogy with single tests, the first kind error rate of a multiple testing procedure is closely related to the number of wrongly rejected tests, that is the number of false positives.
Consider the widespread notation, popularized by Benjamini and Hochberg in [16], and summarized in Table 2.

	Accepted null hypotheses	Rejected null hypotheses	Total
True null hypotheses	U	V	m_0
False null hypotheses	T	S	$m - m_0$
Total	$m - R$	R	m

Table 2 – Number of errors committed when testing m null hypotheses.

In particular, notice that the number of positives, namely R, is an observable random variable whereas the number of false positives V and the number of true null hypotheses m_0 are both unknown.

Unlike for single tests, there are many ways of defining the first kind error rate of a multiple testing procedure. Historically, it was expressed in terms of the *Family Wise Error Rate* (FWER) defined by

$$\text{FWER} = \mathbb{P}(V \geq 1)\,,$$

that is the probability that the procedure makes at least one false positive. In particular, as mentioned in Section 0.6.1, if the multiplicity of the tests is not taken into account and each

single test is performed at level α, then the FWER is controlled by $\alpha \times m$. Indeed, Boole's inequality implies that

$$\text{FWER} \leq \sum_{(\mathcal{H}_{0,i}) \text{ true}} \mathbb{P}(\text{wrongly rejecting } (\mathcal{H}_{0,i})) \leq m_0\alpha \leq m\alpha.$$

Hence, a natural correction, due to Bonferroni (see [87]) consists in dividing each individual level by the number of tests, that is α/m. Yet, when the number of tests is very large, the corrected levels become so small that the procedure may not reject any null hypothesis and then becomes too conservative.

To overcome this difficulty, another notion of first kind error rate has been popularized by Benjamini and Hochberg [16]. The *False Discovery Rate* (FDR) is defined by

$$\text{FDR} = \mathbb{E}\left[\frac{V}{R}\mathbb{1}_{R>0}\right].$$

It represents the expected rate of false positives among the positives. The FDR is weaker than the FWER in the sense that the inequality FDR $\leq$ FWER always holds (with equality when $m_0 = m$). In particular, the FDR leads to procedures which may be less conservative since it is less strict on the number of authorized false positives and it has been the first alternative to the FWER commonly accepted in many fields (such as genomics, clinical studies, interaction networks, etc).

Many multiple testing procedures have been developed in order to control the FWER, the FDR, or even other notions of first kind error rate (see, for instance, the book of Dudoit and van der Laan [47]). Among them lies Benjamini and Hochberg's [16] famous procedure. Let $\{p_1, \ldots, p_m\}$ denote the p-values corresponding to the null hypotheses $\{(\mathcal{H}_{0,1}), \ldots, (\mathcal{H}_{0,m})\}$, and apply the following step-up method.

- First order the p-values: $p_{(1)} \leq \cdots \leq p_{(m)}$, and let $(\mathcal{H}_{0,(i)})$ correspond to $p_{(i)}$.
- Then introduce $\hat{k} = \max\left\{i \ ; \ p_{(i)} \leq \frac{i\alpha}{m}\right\}$.
- Finally, reject the null hypothesis corresponding to the $\hat{k}$ smallest p-values, that is $(\mathcal{H}_{0,(1)}), \ldots, (\mathcal{H}_{0,(\hat{k})})$.

At the root of its popularity are its simplicity, its efficiency to treat all kinds of problematics in many application fields, and its ability to control the FDR under some assumptions. For instance, in [16] is stated that:

> If the p-values are independent, then the Benjamini and Hochberg procedure controls the FDR at level less than or equal to α.

Since the independence assumption is quite restrictive, Benjamini and Yekutiely [17] relaxed it to positive dependencies. More precisely, a random vector $(\xi_1, \ldots, \xi_m)$ is said to satisfy the property of *Positive Regression Dependency on a Subset* I_0 of $\{1, \ldots, m\}$ (PRDS) if, for all increasing set D in $[0,1]^m$ (characterized by "if $x \in D$ and $x \leq y$ then $y \in D$"), and for all i in I_0, the application $u \mapsto \mathbb{P}((p_1, \ldots, p_m) \in D | p_i = u)$ is non-decreasing. In particular, they proved that:

> If the p-values satisfy the PRDS property on the subset I_0 of indices corresponding to the true null hypotheses, then the Benjamini and Hochberg procedure controls the FDR at level less than or equal to α.

Notice that Blanchard and Roquain provide a simpler and attractive proof of this last result as a by-product of a more general result in [23].

0.6.4 The Permutation UE method and application on real data

Recall that the first step in answering the neurobiological problematic consists in testing whether two neurons have independent activities on a time window, and the second step consists in detecting the moments of the experiment where synchronizations occur thanks to multiple testing procedures.

Contribution of this thesis

First step: single tests. Consider a time window $W = [a, b]$ contained in the experiment time interval $[0, T]$. When applying to the neuroscience context either the bootstrap or the permutation tests of independence in the *Linear case* introduced in Section 0.4.3, one first needs to define the function φ in the definition of the test statistic with $h = h_\varphi$ (see (0.4.2) and (0.4.3)). A well-adapted choice for the synchrony analysis is the delayed coincidence count introduced by Tuleau-Malot et al. in [170]. Precisely, the delayed coincidences count, with delay δ, between two point processes X^1 and X^2 (modeling two simultaneously recorded spike trains) on the interval $[a, b]$ is defined by

$$\varphi_\delta^{coinc}(X^1, X^2) = \int_a^b \int_a^b \mathbf{1}_{|u-v|\leq\delta} dN_{X^1}(u) dN_{X^2}(v).$$

It counts the number of times a spike from one neuron and a spike from the other neuron occur at distance less than or equal to δ. Notice that φ_δ^{coinc} cannot be written as a product function and thus well-known bootstrap or permutation tests encountered in the statistical literature, such as Romano's [155] or van der Vaart and Wellner's [173], were not suitable for this problematic.

Instinctively, the upper-tailed bootstrap and permutation tests detect the window W if the number of (delayed) coincidences on W is significantly too large compared to what is expected under independence.

A simulation study has been performed in order to check the validity of both bootstrap and permutation independence tests introduced in Section 0.4.3 based on delayed coincidence count from a practical point of view and is given in Chapter 1, Section 1.5. Moreover, their performance, in terms of first and second kind error rates, is compared to existing classical methods such as the *trial-shuffling* (see [137, 138]), or the single testing procedures of the MTGAUE method (see [170]), on different types of point processes. In a few words, as excepted, the permutation test is the only one to control non-asymptotically its first kind error rate and it seems to perform as well as the other ones in terms of second kind error rate. It is thus the only one considered for the multiple testing step. Moreover, one can see that our bootstrap and permutation tests seem to always perform as well as the finest ones without suffering from the drastic restriction to homogenous Poisson process (as the MTGAUE method) or the lack of centering (as the *trial-shuffling* method).

Second step: multiple tests. The idea, as done, for instance, in [170], is then to cut the experiment time slot $[0, T]$ into several windows W, to simultaneously apply the testing method presented above on each W, and return all the detected windows.

We construct in Chapter 2 a multiple testing method based on Benjamini and Hochberg's procedure which enables to distinguish excitatory synchronizations from inhibitory ones. The idea is to apply simultaneously on each window both upper-tailed and lower-tailed tests.

Hence, given a family of K possibly overlapping windows that covers the whole interval $[0, T]$, we compute both p-values p_W^+ and p_W^- corresponding to the upper-tailed and the lower-tailed permutation tests based on the delayed coincidence count on each W, and apply Benjamini and Hochberg's procedure to the whole family of $m = 2K$ p-values.
At this point, we have almost no theoretical guarantee that this procedure controls the FDR by α. Yet, it seems to perform well from a practical point of view, as shown by the simulation study in Chapter 2, Section 2.4.2.

Application on real data. To complete this study, we applied this procedure to real data, that have been already studied in the literature by Riehle, Grammont and others (see [150, 64, 151, 170]).
The experimental protocol. Some microelectrodes are implanted in the motor cortex of a Rhesus monkey trained to perform a multidirectional pointing task. The animal sits in front of seven touch-sensitive light-emitting diode (LED), one in the center, and six equidistantly placed around it. The experiment starts when the monkey touches the central LED. After a delay of 500ms, one of the peripheral LEDs lights up in green, indicating the target for the upcoming movement; it is the *preparatory signal* (PS). Then, after a delay of either 600ms (with probability 0.3) or 1200ms otherwise, the illuminated peripheral LED turns red, and then becomes a target for the monkey; it is the *response signal* (RS). The monkey was rewarded by a drop of juice at the end of each correct responses, and a resting delay was considered between the trials, justifying the i.i.d. assumption on the data. A more complete description can be found in Chapter 2, Section 2.4.2, together with some specifications about the recording technique.

The obtained results (see Chapter 2, Figure 2.9) show that the new method is able to detect responses in line with the time of the experiment. Moreover, it detects those phenomenon with more guarantee since the procedure does not need stationarity assumptions, and, as a consequence, it should have less false negatives in particular on windows where the data are highly non-stationary.

Organization of this thesis

To summarize, here is a content of the present manuscript.

Chapter 1 contains the construction of both bootstrap and permutation tests of independence. It is devoted to the asymptotic study of these tests in the point process framework. More precisely, the conditional distributions of both bootstrap and permutation test statistics given the observation are shown to converge, with respect to the $\mathbb{L}_2$-Wasserstein metric, to the limit of the null distribution, and this whether the observation does or does not satisfy the null hypothesis. In particular, this proves that both bootstrap and permutation approaches reconstruct the null hypothesis. Then, the asymptotic performance of the corresponding tests is derived. Moreover, the assumptions required for these results are shown to be satisfied by most of the models usually considered in the neuroscience literature. This chapter also includes a simulation study verifying the usability of these new methods in practice, and comparing them to existing classical methods in neuroscience.

Chapter 2 concerns the biological application and aims at a neuroscientific audience. One of the main goals of this chapter, is to convince that bootstrapping non centered statistics may lead to drastic errors. More precisely, if the test statistic is not centered under the null hypothesis, the distribution that is approximated by the bootstrap type approaches is different from the desired null distribution, as illustrated in a simulation study. The other goal of this chapter is to present the methodology of our new testing methods in neuroscience. The p-values of both bootstrap and permutation tests constructed in Chapter 1 and based on the delayed coincidence count are defined, together with the ones of the *trial-shuffling* method, with or without centering. All these procedures based on p-values are compared on simulated data. Then, a Benjamini-Hochberg type multiple testing procedure based on permutation p-values, namely *Permutation Unitary Events method*, is implemented to detect the synchronizations between two spike trains. The practical validity of the method is verified on a simulation study before being applied on real data.

Chapter 3 is dedicated to the study of concentration inequalities for randomly permuted sums. Thanks to Talagrand's fundamental inequalities for random permutations and Ledoux's trick for the transition from the median to the mean, we finally obtain a Bernstein-type concentration inequality. The reasons why such sophisticated tools are needed for the study of the uniform separation rates of the permutation test are also explained.

Chapter 4 is devoted to the non-asymptotic study of the performance in terms of uniform separation rates. In line with the single permutation tests constructed in Chapter 1, an aggregated independence testing procedure, based on a permutation approach and wavelet thresholding, is developed in the density framework. In particular, its uniform separation rate is studied over particular classes of functions, namely weak Besov spaces, with respect to the $\mathbb{L}_2$-metric, and seem to be optimal and adaptive in the minimax sense.

Remaining questions are pointed out in a conclusion chapter, opening new perspectives in theoretical Statistics as well as in neuroscience methodology.
Finally, some mathematical tools are succinctly presented in an Appendix section, recalling some fundamental results used in this thesis.

Chapters 1, 2, and 4 are respectively published, in revision and almost submitted articles, and have therefore been left in their article form. In particular, some considerations and definitions are redundant with this introduction.

Chapter 1

Bootstrap and permutation tests of independence for point processes

Motivated by a neuroscience question about synchrony detection in spike train analysis, we deal with the independence testing problem for point processes. We introduce non-parametric test statistics, which are rescaled general U-statistics, whose corresponding critical values are constructed from bootstrap and randomization/permutation approaches, making as few assumptions as possible on the underlying distribution of the point processes. We derive general consistency results for the bootstrap and for the permutation w.r.t. Wasserstein's metric, which induces weak convergence as well as convergence of second-order moments. The obtained bootstrap or permutation independence tests are thus proved to be asymptotically of the prescribed size, and to be consistent against any reasonable alternative. A simulation study is performed to illustrate the derived theoretical results, and to compare the performance of our new tests with existing ones in the neuroscientific literature.

This Chapter is the fruit of a collaboration with Yann Bouret[1], Magalie Fromont[2] and Patricia Reynaud-Bouret[3]. The corresponding article is published in *The Annals of Statistics*.

Contents

1.1 Introduction 82
1.2 From neuroscience interpretations to general test statistics . . . 85
1.2.1 The testing problem 85
1.2.2 Independence test based on coincidences in neuroscience 85
1.2.3 General U-statistics as independence test statistics 86
1.2.4 Non-degeneracy of the U-statistics under $(\mathcal{H}_0)$ 87
1.3 Bootstrap tests of independence 88
1.3.1 Additional notation: bootstrap and convergence formalism 88
1.3.2 Main assumptions 89
1.3.3 Consistency of the bootstrap approach 91
1.3.4 Convergence of cumulative distribution functions and quantiles . . . 92
1.3.5 Asymptotic properties of the bootstrap tests 94

[1]LPMC, Université Nice Sophia Antipolis
[2]IRMAR, Université Rennes 2
[3]LJAD, Université Nice Sophia Antipolis

1.4 Permutation tests of independence 96
1.4.1 The permutation approach and its known non-asymptotic properties 96
1.4.2 Consistency of the permutation approach 97
1.4.3 Asymptotic properties of the permutation tests 98
1.5 Simulation study 98
1.5.1 Presentation of the study 98
1.5.2 Results 100
1.6 Conclusion 102
1.7 Proofs 103
1.7.1 Proof of Proposition 1.2.1 103
1.7.2 Proof of Proposition 1.3.1 104
1.7.3 Proof of Proposition 1.3.2 105
1.7.4 Proof of Theorem 1.3.1 112
1.7.5 Proof of Proposition 1.3.3 113
1.7.6 Proof of Proposition 1.3.4 114
1.7.7 Proof of Proposition 1.3.5 115
1.7.8 Proof of Corollary 1.3.1 116
1.7.9 Proof of Theorem 1.3.2 117
1.7.10 Proof of Proposition 1.3.6 117
1.7.11 Proof of Theorem 1.4.1 118
1.7.12 Proof of Corollary 1.4.1 136
1.7.13 Proof of Theorem 1.4.2 136
1.8 Additional Results 137
1.8.1 About the non-degeneracy of the U-statistic 137
1.8.2 About the empirical centering assumption 138

1.1 Introduction

Inspired by neuroscience problems, the present work is devoted to independence tests for point processes. The question of testing whether two random variables are independent is of course largely encountered in the statistical literature, as it is one of the central goals of data analysis. From the historical Pearson's chi-square test of independence (see [131, 132]) to the modern test of [66] using kernel methods in the spirit of statistical learning, many non-parametric independence tests have been developed for real-valued random variables or even random vectors. Among them, of particular interest are the tests based on the randomization/permutation principle introduced by Fisher [54], and covered thereafter in the series of papers by Pitman [139, 141], Scheffe [161], Hoeffding [84] for instance, or bootstrap approaches derived from Efron's [50] "naive" one. Note that permutation and bootstrap-based tests have a long history of applications, of which independence tests are just a very small part (see, for instance, [51, 135, 154, 155] for some reviews, or [4, 100, 112, 110, 56] for more recent works). Focusing on independence tests, two families of permutation or bootstrap-based tests may be distinguished at least: the whole family of rank tests including the tests of Hotelling and Pabst [89], Kendall [104], Wolfowitz [179] or Hoeffding [82] on the one hand, the family of Kolmogorov-Smirnov type tests, like Blum, Kiefer and Rosenblatt's [24], Romano's [155] or van der Vaart and Wellner's [173] ones on the other hand.

To describe the properties of these tests, let us recall and fix a few definitions, which are furthermore used throughout this thesis. Tests are said to be *non-parametric* if they are free from the underlying distribution of the observed variables. For any prescribed α in $(0,1)$, tests are said to be *exactly of level* α if their first kind error rate is less than or equal to α whatever the number of observations. This is a non-asymptotic property. Tests are also said to be *asymptotically of size* α if their first kind error rate tends to α when the number of observations tends to infinity. They are said to be *consistent* against some alternative if, under this alternative, their second kind error rate tends to 0 or equivalently their power tends to 1, when the number of observations tends to infinity. Finally, *bootstrap* refers here to bootstrap with replacement. It is thus different from *permutation*, which appears sometimes in the literature as bootstrap without replacement. In this respect, the above mentioned tests of independence are all non-parametric and asymptotically of the prescribed size. Moreover, the tests based on permutation are exactly of the desired level. Some of these tests are proved to be consistent against many alternatives, such as Hoeffding's [82] one and the family of Kolmogorov-Smirnov type tests.

Detecting dependence is also a fundamental old point in the neuroscientific literature (see, e.g., [61]). The neuroscience problem we were initially interested in consists in detecting interactions between occurrences of action potentials on two different neurons simultaneously recorded on n independent trials, as described in [71]. Each recorded set of time occurrences of action potentials for each neuron is usually referred to as a spike train, the spikes being the time occurrences themselves. It is commonly accepted that these spikes are one of the main components of the brain activity (see [166]). So, when observing two spike trains coming from two different neurons, one of the main elementary problem is to assess whether these two spike trains are independent or not. Unfortunately, even if the real recordings of spike trains are discretized in time, and thus belong to finite dimensional spaces, due to the record resolution, the dimension of these spaces is so huge (from ten thousand up to one million) that it is neither realistic nor reasonable to model them by finite dimensional variables, and to apply usual independence tests. Several methods, such as the classical Unitary Events method (see [71] and the references therein), consist in binning the spike trains at first in order to deal with vectorial data with reduced dimension. However, it has been shown that these dimension reduction methods involve an information loss of more than 60% in some cases, making this kind of pre-processing quite proscribed despite its simplicity of use. It is therefore more realistic and reasonable to model recordings of spike trains by finite point processes, and to use independence tests specifically dedicated to such point processes. Asymptotic tests of independence between point processes have already been introduced in [170], but in the particular case of homogeneous Poisson processes. Such a parametric framework is necessarily restrictive and even possibly inappropriate here, as the very existence of any precise underlying distribution for the point processes modeling spike train data is subject to broad debate (see [142, 147]). We thus focus on non-parametric tests of independence for point processes. In this spirit, particular bootstrap methods under the name of *trial-shuffling* have been proposed in [138, 137] for binned data with relatively small dimension, without proper mathematical justification. Besides the loss of information that the binning data pre-processing involves, it appears that the test statistics chosen in these papers do not lead to tests of asymptotic prescribed size as shown by our simulation study.

We here propose to construct new non-parametric tests of independence between two point processes, from the observation of n independent copies of these point processes, with as few assumptions as possible on their underlying distributions. Our test statistics are based on U-statistics (see [164, Chapter 5] for a key reference on U-statistics, or Appendix A.3 for a quick

introduction). The corresponding critical values are obtained from bootstrap or permutation approaches. It has been acknowledged that when both bootstrap and permutation approaches are available, permutation should be preferred, since the corresponding tests are exactly of the desired level [51, p. 218]. Nevertheless, we keep investigating them together, as bootstrap methods — through *trial-shuffling* — are the usual references in neuroscience. Moreover, for specific U-statistics, the corresponding tests share the same properties: both are proved to be asymptotically of the prescribed size and consistent against any reasonable alternative, despite the fact that different tools are used to obtain these results. Indeed, the distance between the bootstrapped distribution and the initial distribution under independence is here directly studied for the bootstrap approach, unlike the permutation approach. Finally, both procedures have good performance in practice when the sample size is moderate to small, as is often the case in neuroscience due to biological or economical reasons.

As U-statistics are usual tools for non-parametric statistical inference, many works deal with the application of bootstrap or permutation to U-statistics. From the original work of Arvesen [8] about the Jackknife of U-statistics, to the recent one of Leucht and Neumann [118], several papers [20, 30, 5, 41] have been devoted to the general problem of bootstrapping a U-statistic. The use of bootstrap or permutation of U-statistics is specially considered in testing problems [88, 36], in particular in dependence detection problems with the Kolmogorov-Smirnov type tests cited above [155, 173].

But all those works exclusively focus on U-statistics of independent and identically distributed (i.i.d.) real-valued random variables or vectors. To our knowledge, there is no previous work on the bootstrap or permutation of general U-statistics for i.i.d. pairs of point processes, as considered in the present paper. The main difficulty thus lies in the nature of the mathematical objects we handle here, that is point processes and their associated point measures which are random measures. The proofs of our results, although inspired by Romano's [153, 155] work and Hoeffding's [84] precursor results on the permutation, are therefore more technical and complex on many aspects detailed in the sequel. In addition, we aim at obtaining the asymptotic distribution of the bootstrapped or permuted test statistics under independence, but also under dependence (see Theorem 1.3.1 and Theorem 1.4.1). Concerning the permutation approach, such a result is, as far as we know, new even for more classical settings than point processes. It thus partially solves a problem stated as open question in [173].

This chapter is organized as follows.

We first present in Section 1.2 the testing problem, and introduce the main notation. Starting from existing works in neuroscience, we introduce our test statistics, based on general kernel-based U-statistics.

Section 1.3 is devoted to our bootstrap approach, and are given new general results about the consistency of the bootstrap for the considered U-statistics, expressed in terms of Wasserstein's metric as in [20]. The convergence is studied under independence as well as under dependence. The corresponding bootstrap independence tests are therefore shown to be asymptotically of the desired size, and consistent against any reasonable alternative. The impact of using Monte Carlo methods to approximate the bootstrap quantiles is also investigated in this section.

Section 1.4 is devoted to the permutation approach which leads, by nature, to non-parametric independence tests exactly of the desired level, and this, even when a Monte Carlo method is used to approximate the permutation quantiles. Are then given new general results about the consistency of the permutation approach when the kernel of the U-statistic has a specific form. These results are still expressed in terms of Wasserstein's metric (see Appendix A.1.2). As a consequence, the corresponding permutation independence tests are proved to satisfy the same asymptotic properties as the bootstrap ones under the null hypothesis as well as under

the same alternatives.
As a comparison of the performance of our tests with existing ones in neuroscience, especially when the sample sizes are moderate or even small, a simulation study is presented in Section 1.5.
A conclusion is given in Section 1.6, and the proofs are detailed in Section 1.7.
Finally, some additional technical results can be found in Section 1.8.

1.2 From neuroscience interpretations to general test statistics

1.2.1 The testing problem

Throughout this chapter, we consider finite point processes defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and observed on $[0,1]$, that is random point processes on $[0,1]$ whose total number of points is almost surely finite (see [38] for instance). Typically, in a neuroscience framework, such finite point processes may represent spike trains recorded on a given finite interval of time, and rescaled so that their values may be assumed to belong to $[0,1]$. The set $\mathcal{X}$ of all their possible values consists of the countable subsets of $[0,1]$. It is equipped with a metric $d_{\mathcal{X}}$ that we introduce in (1.3.3). This metric, issued from the Skorohod topology, makes $\mathcal{X}$ separable and allows to define accordingly Borel sets on $\mathcal{X}$ and by extension on $\mathcal{X}^2$ through the product metric. See Appendix A.2.2 for more details.
The point measure dN_x associated with an element x of $\mathcal{X}$ is defined for all measurable real-valued function f by $\int_{[0,1]} f(u) dN_x(u) = \sum_{u \in x} f(u)$. In particular, the total number of points of x, denoted by $\#x$, is equal to $\int_{[0,1]} dN_x(u)$. Moreover, for a finite point process X defined on $(\Omega, \mathcal{A}, \mathbb{P})$ and observed on $[0,1]$, $\int f(u) dN_X(u)$ becomes a real random variable, defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$.
A pair $X = (X^1, X^2)$ of finite point processes defined on $(\Omega, \mathcal{A}, \mathbb{P})$ and observed on $[0,1]$, has joint distribution P, with marginals P^1 and P^2 if $P(\mathcal{B}) = \mathbb{P}(X \in \mathcal{B})$, $P^1(\mathcal{B}^1) = \mathbb{P}(X^1 \in \mathcal{B}^1)$, and $P^2(\mathcal{B}^2) = \mathbb{P}(X^2 \in \mathcal{B}^2)$, for every Borel set $\mathcal{B}$ of $\mathcal{X}^2$, and all Borel sets $\mathcal{B}^1$, $\mathcal{B}^2$ of $\mathcal{X}$.
Given the observation of an i.i.d. sample $\mathbb{X}_n = (X_1, \dots, X_n)$ from the same distribution P as X, with $X_i = (X_i^1, X_i^2)$ for every $i = 1 \dots n$, we aim at testing $(\mathcal{H}_0)$ "X^1 and X^2 are independent" against $(\mathcal{H}_1)$ "X^1 and X^2 are not independent", which can also be written as

$$(\mathcal{H}_0)\ P = P^1 \otimes P^2 \quad \text{against} \quad (\mathcal{H}_1)\ P \neq P^1 \otimes P^2.$$

1.2.2 Independence test based on coincidences in neuroscience

In the neuroscience issue which initially motivated this work, the sample $\mathbb{X}_n = (X_1, \dots, X_n)$ models pairs of rescaled spike trains issued from two distinct and simultaneously recorded neurons during n trials. Those data are usually recorded on living animals that are repeatedly subject to the same stimulus or that are repeatedly executing the same task. Because there are periods of rest between the records, it is commonly admitted that the n trials are i.i.d. and that the considered i.i.d. sample model is actually realistic. Then, the main dependence feature that needs to be detected between both neurons corresponds to synchronization in time, referred to as coincidences [71]. More precisely, neuroscientists expect to detect if such coincidences occur significantly, that is more than what may be due to chance. They speak in this case of a detected synchrony.

In [170], the notion of coincidence count between two point processes X^1 and X^2 with delay δ ($\delta > 0$) is defined by

$$\varphi_\delta^{coinc}(X^1, X^2) = \int_{[0,1]^2} \mathbb{1}_{|u-v|\leq\delta} dN_{X^1}(u) dN_{X^2}(v) = \sum_{u\in X^1, v\in X^2} \mathbb{1}_{|u-v|\leq\delta}. \tag{1.2.1}$$

Notice that other coincidence count functions have been used in the neuroscience literature such as the binned coincidence count function (that is based on binned data) introduced in [67] or its shifted version [72] (see also [170] for explicit formulas, or Chapter 2, Definition 2.2.1). A further example of possible function used to detect dependence in neuroscience (see [160]) is of the form

$$\varphi^w(X^1, X^2) = \int_{[0,1]^2} w(u, v) dN_{X^1}(u) dN_{X^2}(v). \tag{1.2.2}$$

Under the assumption that both X^1 and X^2 are homogeneous Poisson processes, the independence test of [170] rejects $(\mathcal{H}_0)$ when a test statistic based on $\sum_{i=1}^n \varphi_\delta^{coinc}\left(X_i^1, X_i^2\right)$ is larger than a given critical value. This critical value is deduced from the asymptotic Gaussian distribution of the test statistic under $(\mathcal{H}_0)$. The test is proved to be asymptotically of the desired size, but only under the homogeneous Poisson processes assumption. However, it is now well known that this assumption, as well as many other model assumptions, fails to be satisfied in practice for spike trains (see [142, 147]).

1.2.3 General U-statistics as independence test statistics

In the parametric homogeneous Poisson framework of [170], the expectation of $\varphi_\delta^{coinc}\left(X_i^1, X_i^2\right)$ has a simple expression as a function of δ and the intensities λ_1 and λ_2 of X^1 and X^2. Since λ_1 and λ_2 can be easily estimated, an estimator of this expectation can thus be obtained using the plug-in principle, and subtracted from $\varphi_\delta^{coinc}\left(X_i^1, X_i^2\right)$ to lead to a test statistic with a centered asymptotic distribution under $(\mathcal{H}_0)$.

In the present non-parametric framework where we want to make as few assumptions as possible on the point processes X^1 and X^2, such a centering plug-in tool is not available. We propose to use instead a self-centering trick, which amounts, combined with a rescaling step, to consider the statistic

$$\frac{1}{n(n-1)} \sum_{i\neq i'\in\{1,\ldots,n\}} \left(\varphi_\delta^{coinc}\left(X_i^1, X_i^2\right) - \varphi_\delta^{coinc}\left(X_i^1, X_{i'}^2\right)\right). \tag{1.2.3}$$

It is clear that the function φ_δ^{coinc} used in [170] suits the dependence feature the neuroscientists expect to detect in a spike train analysis. However, it is not necessarily the best choice for other kinds of dependence features to be detected in a general point processes analysis. Note furthermore that the statistic (1.2.3) can be written as a U-statistic of the i.i.d. sample $\mathbb{X}_n = (X_1, \ldots, X_n)$ with a symmetric kernel, as defined by Hoeffding [81].

Let us therefore consider the general independence test statistics which are U-statistics of the form

$$U_{n,h}(\mathbb{X}_n) = \frac{1}{n(n-1)} \sum_{i\neq i'\in\{1,\ldots,n\}} h\left(X_i, X_{i'}\right), \tag{1.2.4}$$

where $h : \mathcal{X}^2 \times \mathcal{X}^2 \to \mathbb{R}$ is a symmetric kernel such that

$(\mathcal{A}_{Cent})$ | For all $n \geq 2$, $U_{n,h}(\mathbb{X}_n)$ is zero mean under $(\mathcal{H}_0)$, that is for X_1 and X_2, i.i.d. with distribution $P^1 \otimes P^2$ on $\mathcal{X}^2$, $\mathbb{E}\left[h\left(X_1, X_2\right)\right] = 0$.

In the sequel, we call *Coincidence case* the case where $h = h_{\varphi_\delta^{coinc}}$, with

$$h_{\varphi_\delta^{coinc}}(x,y) = \frac{1}{2}\left(\varphi_\delta^{coinc}\left(x^1,x^2\right) + \varphi_\delta^{coinc}\left(y^1,y^2\right) - \varphi_\delta^{coinc}\left(x^1,y^2\right) - \varphi_\delta^{coinc}\left(y^1,x^2\right)\right), \quad (1.2.5)$$

so that $U_{n,h_{\varphi_\delta^{coinc}}}(\mathbb{X}_n)$ is equal to the statistic (1.2.3).

A more general choice, which of course includes the above *Coincidence case*, is obtained by replacing φ_δ^{coinc} by any generic integrable function φ. This is the *Linear case*. For any integrable function φ, the kernel h is then taken equal to h_φ, with

$$h_\varphi(x,y) = \frac{1}{2}\left(\varphi\left(x^1,x^2\right) + \varphi\left(y^1,y^2\right) - \varphi\left(x^1,y^2\right) - \varphi\left(y^1,x^2\right)\right). \quad (1.2.6)$$

This example is of utmost importance in the present work since it provides a first proved case of consistency for the permutation approach under the null hypothesis as well as under the alternative (see Theorem 1.4.1). In this case, note that $(\mathcal{A}_{Cent})$ is straightforwardly satisfied, that is $U_{n,h_\varphi}(\mathbb{X}_n)$ is zero mean under $(\mathcal{H}_0)$. Note furthermore that $U_{n,h_\varphi}(\mathbb{X}_n)$ is an unbiased estimator of

$$\int\int \varphi\left(x^1,x^2\right)\left(dP(x^1,x^2) - dP^1(x^1)dP^2(x^2)\right),$$

without any assumption on the underlying point processes. This is therefore a reasonable independence test statistic. If X^1 and X^2 were finite dimensional variables with continuous distributions w.r.t. the Lebesgue measure, this test statistic would be closely related to generalized Kolmogorov-Smirnov tests of independence. For instance, the test statistics of Blum, Kiefer, and Rosenblatt [24], Romano [155], van der Vaart and Wellner in [173] are equivalent to $\sqrt{n}\sup_{v^1\in\mathcal{V}^1,v^2\in\mathcal{V}^2}\left|U_{n,h_{\varphi_{(v^1,v^2)}}}(\mathbb{X}_n)\right|$, where, respectively:

- $\mathcal{V}^1 = \mathcal{V}^2 = \mathbb{R}$ and $\varphi_{(v^1,v^2)}(x^1,x^2) = \mathbb{1}_{]-\infty,v^1]}(x^1)\mathbb{1}_{]-\infty,v^2]}(x^2)$,
- $\mathcal{V}^1$ and $\mathcal{V}^2$ are countable V.-C. classes of subsets of $\mathbb{R}^d$, and $\varphi_{(v^1,v^2)}(x^1,x^2) = \mathbb{1}_{v^1}(x^1)\mathbb{1}_{v^2}(x^2)$,
- $\mathcal{V}^1$ and $\mathcal{V}^2$ are well-chosen classes of real-valued functions, and $\varphi_{(v^1,v^2)}(x^1,x^2) = v^1(x^1)v^2(x^2)$.

Note also the work of [125] based on integrals instead of the supremum of similar quantities with $\varphi_{(v^1,v^2)}(x^1,x^2) = e^{iv^1x^1}e^{iv^2x^2}$. Thus, to our knowledge, the existing test statistics are based on functions φ of product type. However, as seen in Section 1.2.2, when dealing with point processes, natural functions φ, as for instance φ_δ^{coinc}, are not of this type.

1.2.4 Non-degeneracy of the U-statistics under $(\mathcal{H}_0)$

Following the works of Romano [155] or van der Vaart and Wellner [173], the tests we propose here are based on bootstrap and permutation approaches for the above general U-statistics. Most of the assumptions on h depend on the chosen method (permutation or bootstrap) and are postponed to the corresponding section. However, another assumption is common, besides $(\mathcal{A}_{Cent})$, that is

$(\mathcal{A}_{non-deg})$ | For all $n \geq 2$, $U_{n,h}(\mathbb{X}_n)$ is non-degenerate under $(\mathcal{H}_0)$, that is for all X_1 and X_2, i.i.d. with distribution $P^1 \otimes P^2$ on $\mathcal{X}^2$, $\mathrm{Var}(\mathbb{E}\left[h(X_1,X_2)|X_1\right]) \neq 0$.

This assumption is needed in all results with weak convergence to a Gaussian limit, as its variance has to be strictly positive (see, e.g., Proposition 1.3.5 or Theorem 1.4.1). Since under $(\mathcal{H}_0)$, $U_{n,h}(\mathbb{X}_n)$ is assumed to have zero mean, it is degenerate under $(\mathcal{H}_0)$ if and only if for X with distribution $P^1 \otimes P^2$ and for $P^1 \otimes P^2$-almost every x in $\mathcal{X}^2$, $\mathbb{E}[h(x, X)] = 0$.
In the *Linear case*, this condition implies a very particular link between φ and the distribution of the bivariate point process X, which is unknown. The following result gives some basic condition to fulfill $(\mathcal{A}_{non-deg})$ when φ is the coincidence count function.

Proposition 1.2.1. *If the empty set is charged by the marginals, that is if* $P^1(\{\emptyset\}) > 0$ *and* $P^2(\{\emptyset\}) > 0$ *and if* $\varphi_\delta^{coinc}(X_1, X_2)$ *(see (1.2.1)) is not almost surely null under* $(\mathcal{H}_0)$*, then when* h *is given by (1.2.5),* $(\mathcal{A}_{non-deg})$ *is satisfied.*

The proof can be found in the supplementary material together with a more informal discussion on the *Linear case* with $\varphi = \varphi^w$ as given by (1.2.2).
With respect to neuronal data, assuming that the processes may be empty is an obvious assumption as there often exist trials (usually short) where, just by chance, no spikes have been detected. Moreover, practitioners usually choose δ large enough such that coincidences are observed in practice and, therefore, $\varphi_\delta^{coinc}(X_1, X_2)$ is not almost surely null. Hence in practice, the non-degeneracy assumption is always satisfied in the *Coincidence case*.

Throughout this chapter, $(X_i)_i$ denotes a sequence of i.i.d. pairs of point processes, with $X_i = (X_i^1, X_i^2)$ of distribution P on $\mathcal{X}^2$, whose marginals are P^1 and P^2 on $\mathcal{X}$. For $n \geq 2$, let $\mathbb{X}_n = (X_1, \ldots, X_n)$ and $U_{n,h}(\mathbb{X}_n)$ as in (1.2.4), with a fixed measurable symmetric kernel h satisfying $(\mathcal{A}_{Cent})$. To shorten mathematical expression, $U_n(\mathbb{X}_n)$ refers from now on to $U_{n,h}(\mathbb{X}_n)$.

1.3 Bootstrap tests of independence

Since the distribution of the test statistic $U_n(\mathbb{X}_n)$ is not free from the unknown underlying marginal distributions P^1 and P^2 under the null hypothesis $(\mathcal{H}_0)$, we turn to a classical bootstrap approach, which aims at mimicking it, for large, but also moderate or small sample sizes.
To describe this bootstrap approach, and to properly state our results, we give below additional notation, and discuss the main assumptions.

1.3.1 Additional notation: bootstrap and convergence formalism

For j in $\{1, 2\}$, let P_n^j be the empirical marginal distribution defined by

$$P_n^j = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i^j}. \tag{1.3.1}$$

A bootstrap sample from $\mathbb{X}_n$ is denoted by $\mathbb{X}_n^* = \left(X_{n,1}^*, \ldots, X_{n,n}^*\right)$, with $X_{n,i}^* = (X_{n,i}^{*1}, X_{n,i}^{*2})$, and is defined as an n i.i.d. sample from the distribution $P_n^1 \otimes P_n^2$. Then, the bootstrap distribution of interest is the conditional distribution of $\sqrt{n}U_n(\mathbb{X}_n^*)$ given $\mathbb{X}_n$ to be compared with the initial distribution of $\sqrt{n}U_n(\mathbb{X}_n)$ under $(\mathcal{H}_0)$. To state our convergence results as concisely as possible, we use the following classical formalism:

- For any functional $Z : (\mathcal{X}^2)^n \to \mathbb{R}$, $\mathcal{L}(Z, Q)$ denotes the distribution of $Z(\mathbb{Y}_n)$, where $\mathbb{Y}_n$ is an i.i.d. sample from the distribution Q on $\mathcal{X}^2$. In particular, the distribution of $\sqrt{n}U_n(\mathbb{X}_n)$ under $(\mathcal{H}_0)$ is denoted by $\mathcal{L}\left(\sqrt{n}U_n, P^1 \otimes P^2\right)$.

- If the distribution $Q = Q(W)$ depends on a random variable W, $\mathcal{L}(Z, Q|\, W)$ is the conditional distribution of $Z(\mathbb{Y}_n)$, $\mathbb{Y}_n$ being an i.i.d. sample from the distribution $Q = Q(W)$, given W.

  In particular, the conditional distribution of $\sqrt{n}U_n(\mathbb{X}_n^*)$ given $\mathbb{X}_n$ is denoted by $\mathcal{L}\left(\sqrt{n}U_n, P_n^1 \otimes P_n^2 \middle|\, \mathbb{X}_n\right)$.

- " Q-a.s. in $(X_i)_i$" at the end of a statement means that the statement only depends on the sequence $(X_i)_i$, where the X_i's are i.i.d. with distribution Q, and that there exists an event $\mathcal{C}$ only depending on $(X_i)_i$ such that $\mathbb{P}(\mathcal{C}) = 1$, on which the statement is true. Here, Q is usually equal to P.

- "$Q_n \underset{n\to+\infty}{\Longrightarrow} Q$" means that the sequence of distributions $(Q_n)_n$ converges towards Q in the weak sense, that is for any real-valued, continuous and bounded function g, $\int g(z)dQ_n(z) \to_{n\to+\infty} \int g(z)dQ(z)$.

- As usual, $\mathbb{E}^*[\cdot]$ stands for the conditional expectation given $\mathbb{X}_n$.

One of the aims of this work is to prove that the conditional distribution $\mathcal{L}\left(\sqrt{n}U_n, P_n^1 \otimes P_n^2 \middle|\, \mathbb{X}_n\right)$ is asymptotically close to $\mathcal{L}\left(\sqrt{n}U_n, P^1 \otimes P^2\right)$. Following the historical paper by Bickel and Freedman [20], the closeness between these two distributions, which are both distributions on $\mathbb{R}$, is here measured via the $\mathbb{L}^2$-Wasserstein metric (also called Mallows' metric), defined by

$$d_2^2(Q, Q') = \inf\left\{\mathbb{E}\left[(Z - Z')^2\right],\ (Z, Z') \text{ with marginals } Q \text{ and } Q'\right\}, \tag{1.3.2}$$

for all the distributions Q, Q' with finite second-order moments. Recall that convergence w.r.t. d_2 is equivalent to both weak convergence and convergence of second-order moments (see [20, Section 8] or Appendix A.1.2, Proposition A.1.2).

1.3.2 Main assumptions

The random variables we deal with are not real-valued variables but point processes, so the assumptions needed in our results may be difficult to interpret in this setting. We therefore devote this whole section to their description and discussion.
In addition to Assumption $(\mathcal{A}_{Cent})$, we need its following empirical version:

$(\mathcal{A}^*_{Cent})$ | For $x_1 = (x_1^1, x_1^2), \ldots, x_n = (x_n^1, x_n^2)$ in $\mathcal{X}^2$, $\sum_{i_1,i_2,i_1',i_2'=1}^n h\left(\left(x_{i_1}^1, x_{i_2}^2\right), \left(x_{i_1'}^1, x_{i_2'}^2\right)\right) = 0$.

Notice that this assumption, as well as $(\mathcal{A}_{Cent})$, is fulfilled in the *Linear case* where h is of the form h_φ given by (1.2.6), but $(\mathcal{A}^*_{Cent})$ does not imply that h is of the form h_φ (see the supplementary material for a counterexample).

Moment assumptions. Due to the $\mathbb{L}^2$-Wasserstein metric used here to study the consistency of the bootstrap approach, moment assumptions are required. In particular, the variance of $U_n(\mathbb{X}_n)$ should exist, that is,

$(\mathcal{A}_{Mmt})$ | For X_1 and X_2, i.i.d. with distribution P on $\mathcal{X}^2$, $\mathbb{E}\left[h^2(X_1, X_2)\right] < +\infty$,

and more generally we need:

$(\mathcal{A}^*_{Mmt})$ | For X_1, X_2, X_3, X_4 i.i.d. with distribution P on $\mathcal{X}^2$, and for i_1, i_2, i_1', i_2' in $\{1, 2, 3, 4\}$, $\mathbb{E}\left[h^2\left(\left(X_{i_1}^1, X_{i_2}^2\right), \left(X_{i_1'}^1, X_{i_2'}^2\right)\right)\right] < +\infty$.

Notice that when $(\mathcal{A}^*_{Mmt})$ is satisfied, this implies that

- $(\mathcal{A}_{Mmt})$ is satisfied (taking $i_1 = i_2$, $i'_1 = i'_2$, and $i'_1 \neq i_1$),
- for $X \sim P$, $\mathbb{E}\left[h^2(X, X)\right] < +\infty$ (taking $i_1 = i_2 = i'_1 = i'_2$),
- for X_1, X_2 i.i.d. with distribution $P^1 \otimes P^2$, $\mathbb{E}\left[h^2(X_1, X_2)\right] < +\infty$ (taking i_1, i_2, i'_1, i'_2 all different).

A sufficient condition for $(\mathcal{A}^*_{Mmt})$ and $(\mathcal{A}_{Mmt})$ to be satisfied is that there exist positive constants α_1, α_2, C such that for every $x = (x^1, x^2), y = (y^1, y^2)$ in $\mathcal{X}^2$,

$$|h(x,y)| \leq C\left((\#x^1)^{\alpha_1} + (\#y^1)^{\alpha_1}\right)\left((\#x^2)^{\alpha_2} + (\#y^2)^{\alpha_2}\right),$$

with $\mathbb{E}[(\#X^1)^{4\alpha_1}] < +\infty$ and $\mathbb{E}[(\#X^2)^{4\alpha_2}] < +\infty$.

In the *Linear case* where h is of the form h_φ given by (1.2.6), a possible sufficient condition is that there exist some positive constants α_1, α_2, and C such that for every x^1, x^2 in $\mathcal{X}$,

$$|\varphi(x^1, x^2)| \leq C(\#x^1)^{\alpha_1}(\#x^2)^{\alpha_2},$$

with $\mathbb{E}[(\#X^1)^{4\alpha_1}] < +\infty$ and $\mathbb{E}[(\#X^2)^{4\alpha_2}] < +\infty$.

In particular, in the *Coincidence case*, the coincidence count function φ^{coinc}_δ satisfies: for every x^1, x^2 in $\mathcal{X}$, $|\varphi^{coinc}_\delta(x^1, x^2)| \leq (\#x^1)(\#x^2)$. So, $(\mathcal{A}^*_{Mmt})$ and $(\mathcal{A}_{Mmt})$ are satisfied as soon as $\mathbb{E}[(\#X^1)^4] < +\infty$ and $\mathbb{E}[(\#X^2)^4] < +\infty$.

Such moment bounds for the total number of points of the processes are in fact satisfied by many kinds of point processes: discretized point processes at resolution $0 < r < 1$ (see [170] for a definition), which have at most $1/r$ points, Poisson processes, whose total number of points obeys a Poisson distribution having exponential moments of any order, and point processes with bounded conditional intensities, which can be constructed by thinning homogeneous Poisson processes (see [129] or Appendix A.2.1). Similar moment bounds can also be obtained (see [75]) for linear stationary Hawkes processes with positive interaction functions that are classical models in spike train analysis (see, e.g., [134, 170]). This finally may be extended to point processes whose conditional intensities are upper bounded by intensities of linear stationary Hawkes processes with positive interaction functions, by thinning arguments. This includes more general Hawkes processes (see [29]) and in particular Hawkes processes used to model inhibition in spike train analysis (see [75, 170, 148, 147]).

Continuity of the kernel. The set $\mathcal{X}$ can be embedded in the space $\mathcal{D}$ of càdlàg functions on $[0, 1]$ through the identification

$$N : x \in \mathcal{X} \mapsto \left(N_x : t \mapsto \int_0^1 \mathbb{1}_{u \leq t} dN_x(u)\right) \in \mathcal{D}.$$

Notice that the quantity N_x is actually the counting process associated with x (see [28] for instance): at time t, $N_x(t)$ is the number of points of x less than t. Now consider the uniform Skorohod topology on $\mathcal{D}$ (see [22], or Appendix A.2.2), associated with the metric $d_{\mathcal{D}}$ defined by

$$d_{\mathcal{D}}(f, g) = \inf\left\{\varepsilon > 0 \ ; \ \exists \lambda \in \Lambda, \left\{ \begin{array}{l} \sup_{t\in[0,1]} |\lambda(t) - t| \leq \varepsilon, \\ \sup_{t\in[0,1]} |f(\lambda(t)) - g(t)| \leq \varepsilon \end{array} \right\}\right\},$$

where Λ is the set of strictly increasing, continuous mappings of $[0,1]$ onto itself. Notice that here, λ represents a uniformly small deformation of the time scale. Thanks to the identification N above, $\mathcal{X}$ can then be endowed with the topology induced by $d_{\mathcal{X}}$ defined on $\mathcal{X}$ by

$$d_{\mathcal{X}}(x,x') = d_{\mathcal{D}}(N(x),N(x')) \quad \text{for every } x,\ x' \text{ in } \mathcal{X}. \tag{1.3.3}$$

As an illustration, if x and x' are in $\mathcal{X}$, for ε in $(0,1)$, $d_{\mathcal{X}}(x,x') \leq \varepsilon$ implies that x and x' have the same cardinality, and for k in $\{1,\dots,\#x\}$, the k^{th} point of x is at distance less than ε from the kth point of x'. Since $(\mathcal{D}, d_{\mathcal{D}})$ is a separable metric space, so are $(\mathcal{X}, d_{\mathcal{X}})$, $\left(\mathcal{X}^2, d_{\mathcal{X}^2}\right)$, where $d_{\mathcal{X}^2}$ is the product metric defined from $d_{\mathcal{X}}$ (see [46, p. 42], or Appendix A.2.2), and $\left(\mathcal{X}^2 \times \mathcal{X}^2, d\right)$, where d, the product metric defined from $d_{\mathcal{X}^2}$, is given by

$$d\big((x,y),(x',y')\big) = \sup\left\{ \sup_{j=1,2}\left\{d_{\mathcal{X}}(x^j,x'^j)\right\}, \sup_{j=1,2}\left\{d_{\mathcal{X}}(y^j,y'^j)\right\}\right\}, \tag{1.3.4}$$

for every $x=(x^1,x^2)$, $y=(y^1,y^2)$, $x'=(x'^1,x'^2)$, $y'=(y'^1,y'^2)$ in $\mathcal{X}^2$.
The kernel h chosen to define the U-statistic $U_n(\mathbb{X}_n)$ in (1.2.4) should satisfy

$(\mathcal{A}_{Cont})$ | There exists a subset $\mathcal{C}$ of $\mathcal{X}^2 \times \mathcal{X}^2$, such that
(i) h is continuous on $\mathcal{C}$ for the topology induced by d,
(ii) $(P^1 \otimes P^2)^{\otimes 2}(\mathcal{C}) = 1$.

Here are some examples in the *Linear case* for which $(\mathcal{A}_{Cont})$ holds.

Proposition 1.3.1. *Let $w : [0,1]^2 \to \mathbb{R}$ be a continuous integrable function. Then the kernel h_{φ^w} defined on $\mathcal{X}^2 \times \mathcal{X}^2$ by (1.2.2) and (1.2.6) is continuous w.r.t. the topology induced by d, defined by (1.3.4).*

The above result does not apply to $h_{\varphi_\delta^{coinc}}$ but the following one holds.

Proposition 1.3.2. *The coincidence count kernel $h_{\varphi_\delta^{coinc}}$ defined on $\mathcal{X}^2 \times \mathcal{X}^2$ by (1.2.1) and (1.2.6) is continuous w.r.t. the topology induced by d, on*

$$\mathcal{C}_\delta = \left\{ \big((x^1,x^2),(y^1,y^2)\big) \in \mathcal{X}^2 \times \mathcal{X}^2 \,;\, \big(\{x^1\} \cup \{y^1\}\big) \cap \big(\{x^2 \pm \delta\} \cup \{y^2 \pm \delta\}\big) = \emptyset \right\}. \tag{1.3.5}$$

As suggested in [170], when dealing with discretized point processes at resolution r, the right choice for δ is $kr + r/2$ for an integer k, so $(P^1 \otimes P^2)^{\otimes 2}(\mathcal{C}_\delta) = 1$, and $h_{\varphi_\delta^{coinc}}$ satisfies $(\mathcal{A}_{Cont})$. Furthermore, when dealing with independent point processes with conditional intensities, those processes may be constructed by thinning two independent Poisson processes X and X' (see Appendix A.2.1). Hence, in this case, the probability $(P^1 \otimes P^2)^{\otimes 2}$ of $\mathcal{C}_\delta$ in (1.3.5) is larger than $\mathbb{P}(X \cap (X' \pm \delta) = \emptyset)$, whose value is 1. So when dealing with point processes with conditional intensities, $h_{\varphi_\delta^{coinc}}$ also satisfies $(\mathcal{A}_{Cont})$.

1.3.3 Consistency of the bootstrap approach

The validity of the bootstrap approach for our independence tests is due to the following consistency result.

Theorem 1.3.1. *For every $n \geq 2$, let P_n^j for $j=1,2$ be the empirical marginal distributions defined by (1.3.1). Then, under $(\mathcal{A}_{Cent})$, $(\mathcal{A}^*_{Cent})$, $(\mathcal{A}^*_{Mmt})$ and $(\mathcal{A}_{Cont})$,*

$$d_2\left(\mathcal{L}\left(\sqrt{n}U_n, P_n^1 \otimes P_n^2 \middle| \mathbb{X}_n\right), \mathcal{L}\left(\sqrt{n}U_n, P^1 \otimes P^2\right)\right) \underset{n\to+\infty}{\longrightarrow} 0, \ P\text{-a.s. in } (X_i)_i.$$

The proof follows similar arguments to the ones of [20] for the bootstrap of the mean, or to [41] and [118] for the bootstrap of U-statistics. The main novel point here consists in using the identification (1.3.4) and the properties of the separable Skorohod metric space $(\mathcal{D}, d_{\mathcal{D}})$, where weak convergence of sample probability distributions is available (see [174], or Appendix A.1.2, Theorem A.1.4). This theorem derives in fact from the following two propositions which may be useful in various frameworks. The first one states a non-asymptotic result, while the second one gives rather natural results of convergence.

Proposition 1.3.3. *Under $(\mathcal{A}_{Cent})$, $(\mathcal{A}^*_{Cent})$, $(\mathcal{A}^*_{Mmt})$, with the notation of Theorem 1.3.1, there exists an absolute constant $C > 0$ such that for $n \geq 2$,*

$$d_2^2\left(\mathcal{L}\left(\sqrt{n}U_n, P_n^1 \otimes P_n^2 \middle| \mathbb{X}_n\right), \mathcal{L}\left(\sqrt{n}U_n, P^1 \otimes P^2\right)\right)$$
$$\leq C \inf\left\{\mathbb{E}^*\left[\left(h\left(Y^*_{n,a}, Y^*_{n,b}\right) - h\left(Y_a, Y_b\right)\right)^2\right], Y^*_{n,a} \sim P_n^1 \otimes P_n^2,\ Y_a \sim P^1 \otimes P^2,\right.$$
$$\left.\text{and } (Y^*_{n,b}, Y_b) \text{ is an independent copy of } (Y^*_{n,a}, Y_a)\right\}.$$

Comment. In the above proposition, the infimum is taken over all the possible distributions of $(Y^*_{n,a}, Y_a)$ having the correct marginals, $(Y^*_{n,b}, Y_b)$ being just an independent copy of $(Y^*_{n,a}, Y_a)$. In particular, $Y^*_{n,a}$ is not necessarily independent of Y_a.

Proposition 1.3.4. *If $\mathbb{E}\left[|h(X_1, X_2)|\right] < +\infty$, then*

$$U_n(\mathbb{X}_n) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}\left[h(X_1, X_2)\right] = \int h(x, x')dP(x)dP(x'), \quad P\text{-a.s. in } (X_i)_i. \tag{1.3.6}$$

Under $(\mathcal{A}^*_{Mmt})$, one moreover obtains that P-a.s. in $(X_i)_i$,

$$\frac{1}{n^4}\sum_{i,j,k,l=1}^{n} h^2\left(\left(X_i^1, X_j^2\right), \left(X_k^1, X_l^2\right)\right) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}\left[h^2\left(\left(X_1^1, X_2^2\right), \left(X_3^1, X_4^2\right)\right)\right].$$

1.3.4 Convergence of cumulative distribution functions and quantiles

As usual, $\mathcal{N}(m, v)$ stands for the Gaussian distribution with mean m and variance v, $\Phi_{m,v}$ for its cumulative distribution functions (c.d.f.) and $\Phi^{-1}_{m,v}$ for its quantile function. From the results of Rubin and Vitale [158], generalizing Hoeffding's [81] decomposition of non-degenerate U-statistics to the case where the X_i's are not necessarily real-valued random vectors, a central limit theorem for $U_n(\mathbb{X}_n)$ can be easily derived. It is expressed here using the $\mathbb{L}^2$-Wasserstein metric, and is thus slightly stronger than the one stated in equation (1.1) of [90].

Proposition 1.3.5. *Assume that h satisfies $(\mathcal{A}_{non-deg})$, $(\mathcal{A}_{Cent})$, and $(\mathcal{A}_{Mmt})$. Let $\sigma^2_{P^1\otimes P^2}$ be defined by*

$$\sigma^2_{P^1\otimes P^2} = 4\,\mathrm{Var}(\mathbb{E}\left[h\left(X_1, X_2\right)|X_1\right]), \tag{1.3.7}$$

when X_1 and X_2 are $P^1 \otimes P^2$-distributed. Then

$$d_2\left(\mathcal{L}\left(\sqrt{n}U_n, P^1 \otimes P^2\right), \mathcal{N}(0, \sigma^2_{P^1\otimes P^2})\right) \underset{n\to+\infty}{\longrightarrow} 0.$$

Comments.

(*i*) Notice that $(\mathcal{A}_{non-deg})$ is equivalent to $\sigma^2_{P^1\otimes P^2} > 0$. In the case where $(\mathcal{A}_{non-deg})$ does not hold, that is if $\sigma^2_{P^1\otimes P^2} = 0$, the quantity $\sqrt{n}U_n(\mathbb{X}_n)$ tends in probability towards 0. In this case, Theorem 1.3.1 implies that the two distributions $\mathcal{L}\left(\sqrt{n}U_n, P_n^1 \otimes P_n^2 \middle| \mathbb{X}_n\right)$ and $\mathcal{L}\left(\sqrt{n}U_n, P^1 \otimes P^2\right)$ are not only close, but that they are actually both tending to the delta measure concentrated at 0. Indeed, degenerate U-statistics of order 2 have a faster rate of convergence than $\sqrt{n}$ (see [6] for instance for explicit limit theorems). So in this degenerate case, one could not use $\sqrt{n}U_n(\mathbb{X}_n)$ as a test statistic anymore (without changing the normalization). But as mentioned above, $(\mathcal{A}_{non-deg})$ is usually satisfied in practice (see Section 1.2.4 for the *Coincidence case*).

(*ii*) Let us introduce, as in [90], an estimator of $\sigma^2_{P^1\otimes P^2}$, but which is here corrected to be unbiased under $(\mathcal{H}_0)$, namely

$$\hat{\sigma}^2 = \frac{4}{n(n-1)(n-2)} \sum_{i,j,k\in\{1,\dots,n\},\#\{i,j,k\}=3} h(X_i, X_j)h(X_i, X_k),$$

and the statistic

$$S_n = \sqrt{n}U_n(\mathbb{X}_n)/\hat{\sigma}. \tag{1.3.8}$$

From Proposition 1.3.5 combined with Slutsky's lemma (see Appendix A.1, Proposition A.1.1) and the law of large numbers for U-statistics of order 3 (see, e.g., Appendix A.3.2, Theorem A.3.1), one easily derives that under $(\mathcal{H}_0)$, S_n converges in distribution to $\mathcal{N}(0,1)$. This leads to a rather simple but asymptotically satisfactory test: the test which rejects $(\mathcal{H}_0)$ when

$$|S_n| \geq \Phi^{-1}_{0,1}(1-\alpha/2)$$

is indeed asymptotically of size α. It is also consistent against any reasonable alternative P, satisfying $(\mathcal{A}_{Mmt})$ and such that $\mathbb{E}\left[h(X,X')\right] \neq 0$, for X, X' i.i.d. with distribution P. Such a purely asymptotic test may of course suffer from a lack of power when the sample size n is small or even moderate, which is typically the case for the application in neuroscience described in Section 1.2 for biological reasons (from few tens up to few hundreds at best). Though the bootstrap approach is mainly justified by asymptotic arguments, the simulation study presented in Section 1.5 shows its efficiency in a non-asymptotic context, compared to this simpler test.

As Proposition 1.3.5 implies that the limit distribution of $\sqrt{n}U_n(\mathbb{X}_n)$ has a continuous c.d.f., the convergence of the conditional c.d.f. or quantiles of the considered bootstrap distributions holds (see Appendix A.1.2). Note that these conditional bootstrap distributions are discrete, so the corresponding quantile functions are to be understood as the generalized inverses of the cumulative distribution functions.

Corollary 1.3.1. *For $n \geq 2$, with the notation of Theorem 1.3.1, let $\mathbb{X}_n^*$ be a bootstrap sample, that is an i.i.d. n-sample from the distribution $P_n^1 \otimes P_n^2$. Let $\mathbb{X}_n^{\perp\!\!\!\perp}$ be another i.i.d. n-sample from the distribution $P^1 \otimes P^2$ on $\mathcal{X}^2$. Under $(\mathcal{A}_{non-deg})$ and the assumptions of Theorem 1.3.1,*

$$\sup_{z\in\mathbb{R}} \left| \mathbb{P}\left(\sqrt{n}U_n\left(\mathbb{X}_n^*\right) \leq z \middle| \mathbb{X}_n\right) - \mathbb{P}\left(\sqrt{n}U_n(\mathbb{X}_n^{\perp\!\!\!\perp}) \leq z\right) \right| \underset{n\to+\infty}{\longrightarrow} 0, \ P\text{-a.s. in } (X_i)_i.$$

If moreover, for η *in* $(0,1)$, $q^*_{\eta,n}(\mathbb{X}_n)$ *denotes the conditional* η*-quantile of* $\sqrt{n}U_n(\mathbb{X}^*_n)$ *given* $\mathbb{X}_n$ *and* $q^{\perp\!\!\!\perp}_{\eta,n}$ *denotes the* η*-quantile of* $\sqrt{n}U_n(\mathbb{X}^{\perp\!\!\!\perp}_n)$,

$$|q^*_{\eta,n}(\mathbb{X}_n) - q^{\perp\!\!\!\perp}_{\eta,n}| \underset{n\to+\infty}{\longrightarrow} 0, \ P\text{-a.s. in } (X_i)_i. \tag{1.3.9}$$

1.3.5 Asymptotic properties of the bootstrap tests

We are interested in the asymptotic behavior of sequences of tests all based on test statistics of the form $\sqrt{n}U_n(\mathbb{X}_n)$. The bootstrap approach, whose consistency is studied above, allows to define bootstrap-based critical values for these tests. Note that the permutation approach studied in Section 1.4 is based on the same test statistics, but with critical values obtained by permutation. Hence, we introduce here a condensed and common formalism for the upper-, lower- and two-tailed tests considered in this work, taking into account that the only change in our two considered approaches concerns the critical values. This will help to state our results in the shortest manner.

Let α be fixed in $(0,1)$, and q be a sequence of upper and lower critical values of the form

$$q = \left(q^+_{\alpha,n}(\mathbb{X}_n), q^-_{\alpha,n}(\mathbb{X}_n)\right)_{n\geq 2}.$$

From this sequence q, let us now define the family $\Gamma(q)$ of three sequences of tests $\Delta^+ = (\Delta^+_{\alpha,n})_{n\geq 2}$, $\Delta^- = (\Delta^-_{\alpha,n})_{n\geq 2}$, and $\Delta^{+/-} = (\Delta^{+/-}_{\alpha,n})_{n\geq 2}$, where

$$\left\{\begin{array}{lcll} \Delta^+_{\alpha,n}(\mathbb{X}_n) & = & \mathbb{1}_{\sqrt{n}U_n(\mathbb{X}_n) > q^+_{\alpha,n}(\mathbb{X}_n)} & \text{(upper-tailed test)}, \\ \Delta^-_{\alpha,n}(\mathbb{X}_n) & = & \mathbb{1}_{\sqrt{n}U_n(\mathbb{X}_n) < q^-_{\alpha,n}(\mathbb{X}_n)} & \text{(lower-tailed test)}, \\ \Delta^{+/-}_{\alpha,n}(\mathbb{X}_n) & = & \max\left(\Delta^+_{\alpha/2,n}(\mathbb{X}_n), \Delta^-_{\alpha/2,n}(\mathbb{X}_n)\right) & \text{(two-tailed test)}, \end{array}\right. \tag{1.3.10}$$

the last test being implicitly defined by the corresponding choices in $\alpha/2$.

Of course, q, $\Gamma(q)$, as well as Δ^+, Δ^- and $\Delta^{+/-}$, depend on the choice of α, but since α is fixed at the beginning, to keep the notation as simple as possible, this dependence is, like the one in h, omitted in the notation.

Depending on the choice of q, the classical asymptotic properties that can be expected to be satisfied by $\Gamma(q)$ are $(\mathcal{P}_{size})$ and $(\mathcal{P}_{consist.})$ defined by

$$(\mathcal{P}_{size}) \left| \begin{array}{l} \text{Each } \Delta = (\Delta_{\alpha,n})_{n\geq 2} \text{ in } \Gamma(q) \text{ is asymptotically of size } \alpha, \\ \text{that is } \mathbb{P}(\Delta_{\alpha,n}(\mathbb{X}_n) = 1) \underset{n\to+\infty}{\longrightarrow} \alpha \text{ if } P = P^1 \otimes P^2; \end{array}\right.$$

$$(\mathcal{P}_{consist.}) \left| \begin{array}{l} \text{Each } \Delta = (\Delta_{\alpha,n})_{n\geq 2} \text{ in } \Gamma(q) \text{ is consistent,} \\ \text{that is } \mathbb{P}(\Delta_{\alpha,n}(\mathbb{X}_n) = 1) \underset{n\to+\infty}{\longrightarrow} 1, \text{ for every } P \text{ such that} \\ \quad \bullet \ \int h(x,x')dP(x)dP(x') > 0 \text{ if } \Delta = \Delta^+, \\ \quad \bullet \ \int h(x,x')dP(x)dP(x') < 0 \text{ if } \Delta = \Delta^-, \\ \quad \bullet \ \int h(x,x')dP(x)dP(x') \neq 0 \text{ if } \Delta = \Delta^{+/-}. \end{array}\right.$$

Following Corollary 1.3.1, our bootstrap tests family is defined from (1.3.10) by $\Gamma(q^*)$, with

$$q^* = \left(q^*_{1-\alpha,n}(\mathbb{X}_n), q^*_{\alpha,n}(\mathbb{X}_n)\right)_{n\geq 2}. \tag{1.3.11}$$

Theorem 1.3.2. *Let $\Gamma(q^*)$ be the family of tests defined by (1.3.10) and (1.3.11). If $(\mathcal{A}_{non-deg})$, $(\mathcal{A}_{Cent})$, $(\mathcal{A}^*_{Cent})$, $(\mathcal{A}^*_{Mmt})$ and $(\mathcal{A}_{Cont})$ hold, then $\Gamma(q^*)$ satisfies both $(\mathcal{P}_{size})$ and $(\mathcal{P}_{consist.})$.*

Comments. In the *Linear case* where h is equal to h_φ defined by (1.2.6),

$$\int h(x,x')dP(x)dP(x') = \int \varphi(x^1,x^2)\left[dP(x^1,x^2) - dP^1(x^1)dP^2(x^2)\right].$$

This means that under the assumptions of Theorem 1.3.2, the two-tailed test of $\Gamma(q^*)$ is consistent against any alternative such that $\int \varphi(x^1,x^2)dP(x^1,x^2)$ differs from what is expected under $(\mathcal{H}_0)$, that is $\int \varphi(x^1,x^2)dP^1(x^1)dP^2(x^2)$.

(*i*) In particular, in the *Coincidence case* where h is equal to $h_{\varphi_\delta^{coinc}}$ defined by (1.2.5), the assumptions of Theorem 1.3.2 are fulfilled for instance if X^1 and X^2 are discretized at resolution r, with $\delta = kr + r/2$ for some integer k, or if X^1 and X^2 have bounded conditional intensities, with δ large enough so that $\varphi_\delta^{coinc}(X^1,X^2)$ is not a.s. null. Theorem 1.3.2 means in such cases that the corresponding two-tailed test is asymptotically of power 1, for any alternative P such that

$$\int \mathbb{1}_{|v-u|\leq\delta}\mathbb{E}\left[dN_{X^1}(u)dN_{X^2}(v)\right] \neq \int \mathbb{1}_{|v-u|\leq\delta}\mathbb{E}\left[dN_{X^1}(u)\right]\mathbb{E}\left[dN_{X^2}(v)\right].$$

Note that no δ ensuring this condition can be found if heuristically, the repartition of the delays $|v-u|$ between points of X^1 and X^2 is the same under $(\mathcal{H}_0)$ and under $(\mathcal{H}_1)$. For neuroscientists, it means that the cross-correlogram (histogram of the delays, classically represented as a first description of the data) does not show different behaviors in the dependent and independent cases. This would only occur if the dependence could not be measured in terms of delay between points.

(*ii*) Furthermore, when φ is equal to φ^w defined by (1.2.2) with a continuous integrable function w (see Proposition 1.3.1), Theorem 1.3.2 means that the corresponding two-tailed test is consistent against any alternative such that

$$\beta_w = \int w(u,v)\left(\mathbb{E}\left[dN_{X^1}(u)dN_{X^2}(v)\right] - \mathbb{E}\left[dN_{X^1}(u)\right]\mathbb{E}\left[dN_{X^2}(v)\right]\right) \neq 0.$$

For the function w chosen in [160] and under specific Poisson assumptions, β_w is linked to a coefficient in the Haar basis of the so-called interaction function, which measures the dependence between both processes X^1 and X^2. Working non-asymptotically, one of the main results of [160] states, after reformulation in the present setting, that if β_w is larger than an explicit lower-bound, then the second kind error rate of the upper-tailed test is less than a prescribed β in $(0,1)$. Theorem 1.3.2 thus generalizes the result of [160] to a set-up with much less reductive assumptions on the underlying stochastic models, but in an asymptotic way.

Whereas the above family of bootstrap tests $\Gamma(q^*)$ involves an exact computation of the conditional quantiles $q^*_{\eta,n}(\mathbb{X}_n)$, in practice, these quantiles are approximated by a Monte Carlo method. More precisely, let $(B_n)_{n\geq 2}$ be a sequence of possible numbers of Monte Carlo iterations, such that $B_n \to_{n\to+\infty} +\infty$. For $n \geq 1$, let $\left(\mathbb{X}_n^{*1},\ldots,\mathbb{X}_n^{*B_n}\right)$ be B_n independent

bootstrap samples from $\mathbb{X}_n$. Set $\left(U^{*1},\dots,U^{*B_n}\right)=\left(U_n\left(\mathbb{X}_n^{*1}\right),\dots,U_n\left(\mathbb{X}_n^{*B_n}\right)\right)$, and introduce its corresponding order statistic $\left(U^{*(1)},\dots,U^{*(B_n)}\right)$. The considered family of Monte Carlo bootstrap tests is then defined from (1.3.10) by $\Gamma(q^*_{MC})$, with

$$q^*_{MC}=\left(\sqrt{n}U^{*(\lceil(1-\alpha)B_n\rceil)},\sqrt{n}U^{*(\lfloor\alpha B_n\rfloor+1)}\right)_{n\geq 2}. \tag{1.3.12}$$

Proposition 1.3.6. *Let $\Gamma(q^*_{MC})$ be the family of Monte Carlo bootstrap tests defined by (1.3.10) and q^*_{MC} in (1.3.12). Under the same assumptions as in Theorem 1.3.2, then $\Gamma(q^*_{MC})$ also satisfies both $(\mathcal{P}_{size})$ and $(\mathcal{P}_{consist.})$.*

1.4 Permutation tests of independence

1.4.1 The permutation approach and its known non-asymptotic properties

Consider a random permutation Π_n, uniformly distributed on the set $\mathfrak{S}_n$ of permutations of $\{1,\dots,n\}$, and independent of $\mathbb{X}_n$. Then a permuted sample from $\mathbb{X}_n$ is defined by $\mathbb{X}_n^{\Pi_n}=\left(X_1^{\Pi_n},\dots,X_n^{\Pi_n}\right)$ with $X_i^{\Pi_n}=\left(X_i^1,X^2_{\Pi_n(i)}\right)$. In the same formalism as for the bootstrap approach, for $n\geq 2$ and η in $(0,1)$, let $q^\star_{\eta,n}(\mathbb{X}_n)$ denote the η-quantile of $\mathcal{L}\left(\sqrt{n}U_n,P_n^\star|\mathbb{X}_n\right)$, where $P_n^\star$ stands for the conditional distribution of $\mathbb{X}_n^{\Pi_n}$ given $\mathbb{X}_n$. The family of permutation tests is then defined by $\Gamma\left(q^\star\right)$ (see (1.3.10)), with

$$q^\star=\left(q^\star_{1-\alpha,n}\left(\mathbb{X}_n\right),q^\star_{\alpha,n}\left(\mathbb{X}_n\right)\right)_{n\geq 2}. \tag{1.4.1}$$

As for the bootstrap approach, in practice, the sequence of quantiles $q^\star$ is approximated by a Monte Carlo method. So, let $(B_n)_{n\geq 2}$ be a sequence of numbers of Monte Carlo iterations, such that $B_n\to_{n\to+\infty}+\infty$. For $n\geq 1$, let $\left(\Pi_n^1,\dots,\Pi_n^{B_n}\right)$ be a sample of B_n i.i.d. random permutations uniformly distributed on $\mathfrak{S}_n$. Set

$$\left(U^{\star 1},\dots,U^{\star B_n}\right)=\left(U_n\left(\mathbb{X}_n^{\Pi_n^1}\right),\dots,U_n\left(\mathbb{X}_n^{\Pi_n^{B_n}}\right)\right),$$

and denote by $U^{\star B_n+1}=U_n(\mathbb{X}_n)$ the U-statistic computed on the original sample $\mathbb{X}_n$. The order statistic associated with $\left(U^{\star 1},\dots,U^{\star B_n+1}\right)$ is denoted as usual by $\left(U^{\star(1)},\dots,U^{\star(B_n+1)}\right)$. The considered family of Monte Carlo permutation tests is then defined from (1.3.10) by $\Gamma\left(q^\star_{MC}\right)$, with

$$q^\star_{MC}=\left(\sqrt{n}U^{\star(\lceil(1-\alpha)(B_n+1)\rceil)},\sqrt{n}U^{\star(\lfloor\alpha(B_n+1)\rfloor+1)}\right)_{n\geq 2}. \tag{1.4.2}$$

The main advantage of the above families of permutation tests is that any test $\Delta_{\alpha,n}$ from either $\Gamma(q^\star)$ or $\Gamma(q^\star_{MC})$ is exactly of the desired level α, that is,

$$\text{if } P=P^1\otimes P^2,\quad \mathbb{P}(\Delta_{\alpha,n}(\mathbb{X}_n)=1)\leq\alpha. \tag{1.4.3}$$

Such non-asymptotic results for the permutation tests are well known (see, for instance, [156, Lemma 1] and [135]). Though similar results are since recently available for bootstrap tests in other settings [48, 7, 58], there is no known exact counterpart for the bootstrap in the present context.

1.4.2 Consistency of the permutation approach

In this section, we focus on the *Linear case* where h is of the form h_φ for some integrable function φ, as defined in (1.2.6). Indeed, it is the most general case for which we are able to prove a combinatorial central limit theorem under any alternative as well as under the null hypothesis (Theorem 1.4.1). Hence in this section, U_n refers to U_{n,h_φ}. Notice that the centering assumption $(\mathcal{A}_{Cent})$ is then always satisfied by $U_n(\mathbb{X}_n)$. We here only need the following moment assumption:

$(\mathcal{A}_{\varphi,Mmt})$ | For (X^1, X^2) with distribution P or $P^1 \otimes P^2$ on $\mathcal{X}^2$, $\mathbb{E}\left[\varphi^4(X^1, X^2)\right] < \infty$.

Though we have no exact counterpart of Theorem 1.3.1 for our permutation approach, the following result combined with Proposition 1.3.5 gives a similar result.

Theorem 1.4.1. *For all $n \geq 2$, let $P_n^\star$ be the conditional distribution of a permuted sample given $\mathbb{X}_n$. In the* Linear case *where the kernel h is of the form (1.2.6) for an integrable function φ, under $(\mathcal{A}_{non-deg})$ and $(\mathcal{A}_{\varphi,Mmt})$, with the notation of Section 1.3,*

$$d_2\left(\mathcal{L}\left(\sqrt{n}U_n, P_n^\star \middle| \mathbb{X}_n\right), \mathcal{N}\left(0, \sigma^2_{P^1\otimes P^2}\right)\right) \xrightarrow[n\to+\infty]{\mathbb{P}} 0, \tag{1.4.4}$$

where $\xrightarrow{\mathbb{P}}$ stands for the usual convergence in $\mathbb{P}$-probability.

Comments. As pointed out above, unlike the bootstrap approach, the conditional permutation distribution of the test statistic is not here directly compared to the initial distribution of the test statistic under the null hypothesis. It is in fact compared to the Gaussian limit distribution of the test statistic under the null hypothesis, when the non-degeneracy assumption $(\mathcal{A}_{non-deg})$ holds. Moreover, the convergence occurs here in probability and not almost surely, but note that no continuity assumption for the kernel h_φ is used anymore. The price to pay is that the moment assumption is stronger than the one used for the bootstrap. This assumption, due to our choice to use an existing central limit theorem for martingale difference arrays in the proof, is probably merely technical and maybe dispensable. Indeed, the result of Theorem 1.4.1 is close to asymptotic results for permutation known as *combinatorial central limit theorems* (see [83, 135], or the introduction of Chapter 3), where this kind of higher moment assumption can be replaced by some Lindeberg conditions [74, 127, 80]. However, all these existing results can only be applied directly in our case either when $(X_i)_i$ is deterministic or under the null hypothesis. To our knowledge, no combinatorial central limit theorem has been proved for non-deterministic and non-exchangeable variables, like here under any alternative.

The above result is thus one of the newest results presented here and its scope is well beyond the only generalization to the point processes setting. Indeed, because it holds not only under $(\mathcal{H}_0)$ but also under $(\mathcal{H}_1)$, it goes further than any existing one for independence test statistics such as the ones of Romano [155]. The behavior under $(\mathcal{H}_1)$ of the permuted test statistic of van der Vaart and Wellner was also left as an open question in [173].

From Theorem 1.4.1, we deduce the following corollary.

Corollary 1.4.1. *Under the assumptions of Theorem 1.4.1 and with the notation of Proposition 1.3.5, for η in $(0,1)$,*

$$q^\star_{\eta,n}(\mathbb{X}_n) \xrightarrow[n\to+\infty]{\mathbb{P}} \Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(\eta).$$

1.4.3 Asymptotic properties of the permutation tests

As for the bootstrap tests, we obtain the following result.

Theorem 1.4.2. *Let $\Gamma(q^\star)$ and $\Gamma(q^\star_{MC})$ be the families of permutation and Monte Carlo permutation tests defined by (1.3.10) combined with (1.4.1) and (1.4.2) respectively. In the* Linear case, *if $(\mathcal{A}_{non-deg})$ and $(\mathcal{A}_{\varphi,Mmt})$ hold, then $\Gamma(q^\star)$ and $\Gamma(q^\star_{MC})$ both satisfy $(\mathcal{P}_{size})$ and $(\mathcal{P}_{consist.})$.*

1.5 Simulation study

In this section, we study our testing procedures from a practical point of view, by giving estimations of the size and the power for various underlying distributions that are coherent with real neuronal data. This allows to verify the usability of these new methods in practice, and to compare them with existing classical methods. A real data sets study and a more operational and complete method for neuroscientists derived from the present ones is the subject of Chapter 2. The programs have been optimized, parallelized in `C++` and interfaced with `R`. The code is available at `https://github.com/ybouret/neuro-stat`.

1.5.1 Presentation of the study

All along the study, h is taken equal to $h_{\varphi_\delta^{coinc}}$ (see (1.2.5)), where φ_δ^{coinc} is defined in (1.2.1) and $\alpha = 0.05$. We only present the results for upper-tailed tests, but an analogous study has been performed for lower-tailed tests with similar results. Five different testing procedures are compared.

Testing procedures

(`CLT`) Test based on the central limit theorem for U-statistics (see Proposition 1.3.5) which rejects $(\mathcal{H}_0)$ when the test statistic S_n in (1.3.8) is larger than the $(1-\alpha)$-quantile of the standard normal distribution.

(`B`) Monte Carlo bootstrap upper-tailed test of $\Gamma(q^*_{MC})$ defined by (1.3.10) and (1.3.12).

(`P`) Monte Carlo permutation upper-tailed test of $\Gamma(q^\star_{MC})$ defined by (1.3.10) and (1.4.2).

(`GA`) Upper-tailed tests introduced in [170, Definition 3] under the notation $\Delta^+_{GAUE}(\alpha)$, based on a Gaussian approximation of the total number of coincidences.

(`TS`) Trial-shuffling test based on a Monte Carlo approximation of the p-value introduced in [137, equation (3)], but adapted to the present notion of coincidences. This test is the reference distribution free method for neuroscientists.

More precisely, let $C(\mathbb{X}_n) = \sum_{i=1}^n \varphi_\delta^{coinc}\left(X_i^1, X_i^2\right)$ be the total number of coincidences of $\mathbb{X}_n$. Then, the trial-shuffling method consists in uniformly drawing with replacement n i.i.d. pairs of indices $\{(i^*(k), j^*(k))\}_{1\le k\le n}$ in $\{(i,j), 1\le i\neq j\le n\}$, and considering the associated TS-sample $\mathbb{X}_n^{TS} = ((X^1_{i^*(k)}, X^2_{j^*(k)}))_{1\le k\le n}$. The Monte Carlo p-value is defined by

$$\alpha_B^{TS} = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}_{C\left(\mathbb{X}_n^{TS,b}\right)\ge C(\mathbb{X}_n)},$$

where $\mathbb{X}_n^{TS,1}, \dots, \mathbb{X}_n^{TS,B}$ are B independent TS-samples, and the test rejects $(\mathcal{H}_0)$ if $\alpha_B^{TS} \leq \alpha$. This procedure is therefore close in spirit to our bootstrap procedure except that it is applied on a non-centered quantity under $(\mathcal{H}_0)$, namely $C(\mathbb{X}_n)$.

The number B of steps in the Monte Carlo methods is taken equal to 10000.

Simulated data

Various types of point processes are simulated here to check the distribution free character of our approaches and to investigate their limits. Of course, each of the considered point processes satisfies the moment assumptions on the number of points so that the theorems in this chapter can be applied. From now on and to be coherent with the neuroscience application which originally motivated this work, the point processes are simulated on $[0, 0.1]$. Indeed the following experiments have been done to match neurophysiological parameters [170, 67] and the classical necessary window for detection is usually of the duration of 0.1 seconds.

Estimation of the size. The three data sets simulated under $(\mathcal{H}_0)$ consist of i.i.d. samples of pairs of independent point processes. For simplicity, both processes have the same distribution, though this is not required.

Exp. A Homogeneous Poisson processes on $[0, 0.1]$ with intensity $\lambda = 60$.

Exp. B Inhomogeneous Poisson processes with intensity $f_\lambda : t \in [0, 0.1] \mapsto \lambda t$ and $\lambda = 60$.

Exp. C Hawkes processes as detailed in [170], that is point processes with conditional intensity $\lambda(t) = \max\left\{0, \mu - \int_0^t \nu \, \mathbb{1}_{]0,r]}(t-s)\, dN_X(s)\right\}$, for t in $[0, 0.1]$, with spontaneous intensity $\mu = 60$, refractory period $r = 0.001$, and $\nu > \mu$ such that for all point T in X and t in $]T, T + r]$, $\lambda(t) = 0$. This choice of ν prevents two points to occur at a distance less than the refractory period r to reflect typical neuronal behavior. This model is also sometimes called Poisson process with dead time.

Study of the power. The three data sets simulated under $(\mathcal{H}_1)$ are such that the number of coincidences is larger than expected under $(\mathcal{H}_0)$. The models (injection or Hawkes) are classical in neuroscience and already used in [170, 71].

Exp. D Homogeneous injection model. $X^1 = X^1_{ind} \cup X_{com}$ and $X^2 = X^2_{ind} \cup X_{com}$, X^1_{ind} and X^2_{ind} being two independent homogeneous Poisson processes with intensity $\lambda_{ind} = 54$, X_{com} being a common homogeneous Poisson process with intensity $\lambda_{com} = 6$, independent of X^1_{ind} and X^2_{ind}.

Exp. E Inhomogeneous injection model. Similar to **Exp. D**, X^1_{ind} and X^2_{ind} being two independent inhomogeneous Poisson processes with intensity $f_{\lambda_{ind}}$ (see **Exp. B**), $\lambda_{ind} = 54$, X_{com} being a homogeneous Poisson process with intensity $\lambda_{com} = 6$, independent of X^1_{ind} and X^2_{ind}.

Exp. F Dependent bivariate Hawkes processes. The coordinates X^1 and X^2 of a same pair respectively have the conditional intensities

$$\begin{cases} \lambda^1(t) = \max\left\{0, \mu - \int_0^t \nu \mathbb{1}_{]0,r]}(t-s)\, dN_{X^1}(s) + \int_0^t \eta \mathbb{1}_{]0,u]}(t-s)\, dN_{X^2}(s)\right\}, \\ \lambda^2(t) = \max\left\{0, \mu - \int_0^t \nu \mathbb{1}_{]0,r]}(t-s)\, dN_{X^2}(s) + \int_0^t \eta \mathbb{1}_{]0,u]}(t-s)\, dN_{X^1}(s)\right\}, \end{cases} \tag{1.5.1}$$

with the spontaneous intensity $\mu = 54$, the interaction intensity $\eta = 6$ in the period designated by $u = 0.005$ and the refractory period designated by $r = 0.001$ with $\nu \gg \mu + \eta u$ such that once again, $\lambda^j(t)$ is null on each $]T, T+r]$, for T in X^j. We arbitrarily took $\nu = 50(2\mu + \eta)$.

1.5.2 Results

Varying number of trials n. In Figure 1.1, the delay is fixed at $\delta = 0.01$ and the number n of trials varies in $\{10, 20, 50, 100\}$. Note that when the number of trials is too small ($n = 10$), the estimated variance in (CLT) is sometimes negative, therefore, the test cannot be implemented.
The left-hand side of Figure 1.1 corresponds to estimated sizes. On the one hand, one can see in the case of homogeneous Poisson processes (**Exp. A**) and in the case of refractory Hawkes processes (**Exp. C**) that the methods (CLT), (B), (P) and (GA) are quite equivalent, but the size (first kind error rate) seems less controlled in the bootstrap approach (B) especially for small numbers of trials. Yet, one can see the convergence of the size of the bootstrap test towards α as the number of trials tends to infinity, which illustrates Proposition 1.3.6. Note that the (CLT) test also has a well controlled size even if it cannot be used for very small n. On the other hand, in the case of inhomogeneous Poisson processes (**Exp. B**), one can see that the (GA) test has a huge size and is thus inadequate here. Indeed, it is based on the strong assumption that the data are homogeneous Poisson processes though they are in fact strongly non-stationary. The test tends thus to reject the independence null hypothesis even when the data are independent. Finally, in the three considered cases, the (TS) approach has a very small size, and is thus too conservative as one can see in the power study. The study of Chapter 2 shows that this lack of performance is due to the fact that the (TS) approach is applied here on a quantity which is not correctly centered.
The right-hand side of Figure 1.1 corresponds to estimated powers, which increase as n grows. This is in line with the consistency of the tests. Now, as it could be expected when looking at its estimated sizes, for the (TS) approach, the estimated powers are distinctly lower than the ones for the other methods, which confirms its conservative behavior. The other approaches are more similar in **Exp. D** or **Exp. F** though (B) clearly seems to outperform all tests, but at the price of a less controlled size. Note that in the inhomogeneous case (**Exp. E**), (GA) seems to have the best power, but this time, at the price of a totally uncontrolled size.
This part of the simulation study illustrates the convergences of the size and the power of the bootstrap and permutation tests introduced here. The permutation approach seems to actually guarantee the best control of the size as expected, as compared with the bootstrap approach. Nevertheless, both approaches are quite effective for any considered kind of point processes and any sample size, unlike the (GA) test which has very restrictive assumptions. The reference method (TS) for neuroscientists is clearly too conservative. Moreover, the (CLT) test seems to have also satisfying results, but with a slower convergence than the (B) and (P) tests. This seems to illustrate that the conditional bootstrap and permutation distributions give better approximations of the original one under independence than a simple central limit theorem. This phenomenon is well known as the second-order accuracy of the bootstrap in more classical frameworks.

Varying delay δ. We now investigate the impact of the choice for the delay δ by making δ vary in $\{0.001, 0.005, 0.01, 0.02\}$ for a fixed number of trials $n = 50$. The results for the sizes being similar to the previous study, only the estimated powers are presented in Figure 1.2. On

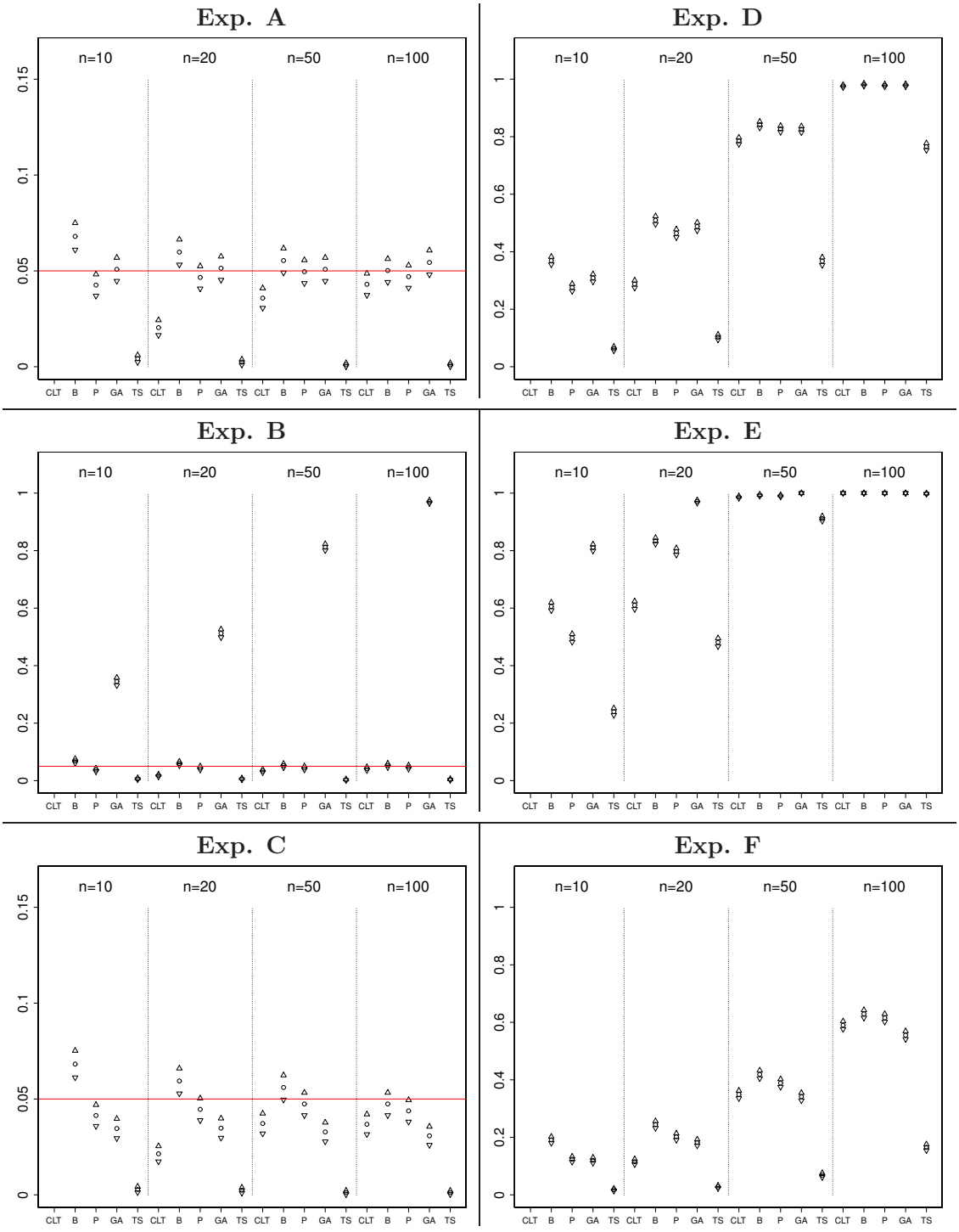


Figure 1.1 – Estimated sizes and powers for various numbers of trials n, all the tests being performed with a level $\alpha = 0.05$. The circles represent the percentage of rejection on 5000 simulations for each method, the triangles represent the corresponding endpoints of a 95% confidence interval. The corresponding experiments are described in Section 1.5.1.

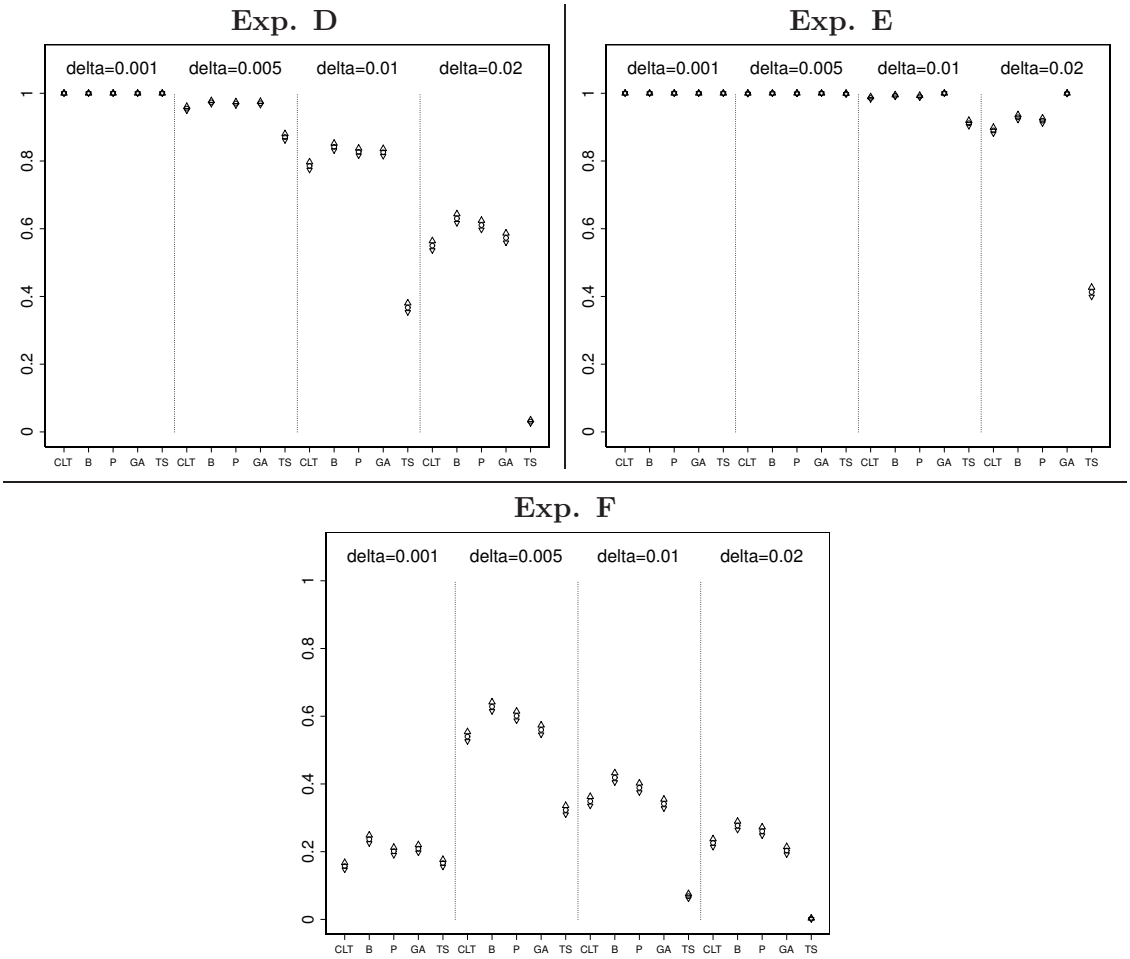


Figure 1.2 – Estimated powers for different δ. Same convention as in Figure 1.1.

the top row of Figure 1.2, the same process is injected in both coordinates: the coincidences are exact in the sense that they have no delay. Therefore, the best choice for the delay parameter δ is the smallest possible value: the obtained power is 1 for very small δ's (e.g., $\delta = 0.001$) and then decreases as δ increases. On the contrary on the bottom row, it can be noticed that the highest power is for $\delta = 0.005$ which is the exact length of the interaction period u in (1.5.1). Once again, the (`TS`) method performs poorly, as does the (`CLT`) method. The three other methods seem to be quite equivalent except in the inhomogeneous case (**Exp. E**) where the (`GA`) method has a power always equal to 1, but at the price of an uncontrolled size.

1.6 Conclusion

In the present paper, we have introduced non-parametric independence tests between point processes based on U-statistics. The proposed critical values are obtained either by bootstrap or permutation approaches. We have shown that both methods share the same asymptotic properties under the null hypothesis as well as under the alternative. From a theoretical point of view, the main asymptotic results (Theorem 1.3.1 and Theorem 1.4.1) have almost the same

flavor. However, there are additional assumptions in the permutation case which make the bootstrap results more general (despite the additional continuity assumption, which is very mild). From a more concrete point of view, it is acknowledged (see, e.g., [51]) that permutation should be preferred because of its very general non-asymptotic properties (1.4.3). This is confirmed by the experimental study, where clearly permutation leads to a better first kind error rate control. However, both approaches perform much better than a naive procedure, based on a basic application of a central limit theorem, when the number of observation is small. They also outperform existing procedures of the neuroscience literature, namely [170], which assumes the point processes to be homogeneous Poisson processes and the trial-shuffling procedures [138, 137], which are biased bootstrap variants applied on a non-centered quantity. One of the main open questions with respect to the existing literature is whether our results can be extended to test statistics as $\sup_h U_{n,h}$. A first obstacle to this question lies in the nature of the observed random variables (point processes) and the fact that controlling such a supremum leads to controlling the whole U-process. This difficulty can probably be overcome, since the asymptotic Gaussian behavior of similar statistics has already been proved in general spaces under $(\mathcal{H}_0)$ for product type kernels (see [27]). The study of such behavior under $(\mathcal{H}_1)$ is surely much more complex. A second obstacle comes from a more practical aspect. In neuroscience, and in the particular case of coincidence count, the use of $\sup_\delta U_{n,h_{\varphi_\delta^{coinc}}}$ leads to the following fundamental problems. On the one hand, such a statistic may not be computable if δ varies in a too large space, typically $[0,1]$. On the other (more important) hand, neuroscientists are especially interested in the value of δ which leads to a rejection, since it actually provides the delay of interaction (see also Section 1.5). In this respect, our work in Chapter 2 involves multiple testing aspects, which may answer this issue.

1.7 Proofs

All along this section, C and C' denote positive constants, that may vary from one line to another one.

1.7.1 Proof of Proposition 1.2.1

We focus on the *Coincidence case*. According to the comment following the definition of $(\mathcal{A}_{non-deg})$, $U_n(\mathbb{X}_n)$ is non-degenerate under $(\mathcal{H}_0)$ if one can find some Borel set $\mathcal{B}$ of $\mathcal{X}^2$ such that $P^1 \otimes P^2(\mathcal{B}) > 0$ and such that for all x in $\mathcal{B}$, $\mathbb{E}\left[h_{\varphi_\delta^{coinc}}(x, X)\right] \neq 0$, where X has distribution $P^1 \otimes P^2$.
Consider $\mathcal{B} = \{(\emptyset, \emptyset)\}$. Then $P^1 \otimes P^2(\mathcal{B}) = P^1(\{\emptyset\}) P^2(\{\emptyset\}) > 0$.
Moreover, as $\varphi_\delta^{coinc}(\cdot, \emptyset)$ and $\varphi_\delta^{coinc}(\emptyset, \cdot)$ are both the zero function, under $(\mathcal{H}_0)$,

$$\begin{aligned}\mathbb{E}\left[h_{\varphi_\delta^{coinc}}((\emptyset,\emptyset), X)\right] &= \frac{1}{2}\left(\mathbb{E}\left[\varphi_\delta^{coinc}(\emptyset,\emptyset) + \varphi_\delta^{coinc}(X^1, X^2) - \varphi_\delta^{coinc}(\emptyset, X^2) - \varphi_\delta^{coinc}(X^1, \emptyset)\right]\right) \\ &= \frac{1}{2}\mathbb{E}\left[\varphi_\delta^{coinc}\left(X^1, X^2\right)\right] > 0,\end{aligned}$$

as $\varphi_\delta^{coinc}\left(X^1, X^2\right)$ is non-negative and not almost surely null under $(\mathcal{H}_0)$.

See also appendix 1.8.1 for further results on the non-degeneracy of the U-statistic in more general cases.

1.7.2 Proof of Proposition 1.3.1

Consider $w : [0,1]^2 \to \mathbb{R}$ a continuous integrable function. Let us prove that $h = h_{\varphi^w}$ given by (1.2.2) and (1.2.6) is continuous for the topology induced by d (see (1.3.4)). Recall that for $x_1 = (x_1^1, x_1^2)$ and $x_2 = (x_2^1, x_2^2)$ in $(\mathcal{X}^2)$,

$$h_{\varphi^w}(x_1, x_2) = \frac{1}{2}\left(\varphi^w(x_1^1, x_1^2) + \varphi^w(x_2^1, x_2^2) - \varphi^w(x_1^1, x_2^2) - \varphi^w(x_2^1, x_1^2)\right).$$

The first step is to show that for each i, j in $\{1, 2\}$, the projection defined by

$$p_{i,j} : \begin{pmatrix} ((\mathcal{X}^2)^2, d) & \longrightarrow & (\mathcal{X}^2, d_{\mathcal{X}^2}) \\ \left((x_1^1, x_1^2), (x_2^1, x_2^2)\right) & \longmapsto & \left(x_i^1, x_j^2\right) \end{pmatrix},$$

is continuous. Let $\mathbf{x} = \left((x_1^2, x_1^2), (x_2^1, x_2^2)\right)$ and $\mathbf{x}' = \left(\left(x'^1_1, {}'^2_1\right), \left(x'^1_2, x'^2_2\right)\right)$ in $(\mathcal{X}^2)^2$. Then,

$$d_{\mathcal{X}^2}\left(p_{i,j}(\mathbf{x}), p_{i,j}(\mathbf{x}')\right) = d_{\mathcal{X}^2}\left(\left(x_i^1, x_j^2\right), \left(x'^1_i, x'^2_j\right)\right) \leq d\left(\mathbf{x}, \mathbf{x}'\right).$$

Hence, $p_{i,j}$ is 1-Lipschitz and therefore continuous.

The second step is to show that if w is continuous on $\left([0,1]^2, \|\cdot\|_\infty\right)$, with $\|(u,v)-(u',v')\|_\infty = \max\{|u-u'|, |v-v'|\}$, then φ^w is also continuous.

Let $\varepsilon > 0$ and for z in $\mathcal{X}$, recall that N_z is the counting process associated with z, defined by

$$N_z(t) = \int_0^1 \mathbb{1}_{u \leq t} dN_z(u).$$

First notice that, w being continuous on the compact set $[0,1]^2$, w is uniformly continuous. Thus one can find some η in $(0,1)$ such that, for all (u,v), (u',v') in $[0,1]^2$,

$$\|(u,v) - (u',v')\|_\infty \leq \eta \quad \text{implies} \quad \left|w(u,v) - w(u',v')\right| \leq \varepsilon. \tag{1.7.1}$$

Consider such η.

Let $\{x_n\}_{n\geq 0}$ be a sequence in $\mathcal{X}^2$ such that $d_{\mathcal{X}^2}(x_n, x_0) \underset{n\to+\infty}{\longrightarrow} 0$ and let us show that

$$\varphi^w(x_n) \underset{n\to+\infty}{\longrightarrow} \varphi^w(x_0).$$

There exists n_0 in $\mathbb{N}$ such that for all $n \geq n_0$, $d_{\mathcal{X}^2}(x_n, x_0) \leq \eta$. Then, for such n, by definition of $d_{\mathcal{X}^2}$, we have that $d_{\mathcal{D}}\left(N_{x_n^1}, N_{x_0^1}\right) \leq \eta$ and $d_{\mathcal{D}}\left(N_{x_n^2}, N_{x_0^2}\right) \leq \eta$. Thus, by definition of $d_{\mathcal{D}}$,

$$\exists \lambda_n^1 \in \Lambda \;/\; \begin{cases} \sup_{t\in[0,1]} \left|\lambda_n^1(t) - t\right| \leq \eta, & (1\text{-}i) \\ \sup_{t\in[0,1]} \left|N_{x_n^1}(t) - N_{x_0^1}\left(\lambda_n^1(t)\right)\right| \leq \eta, & (1\text{-}ii) \end{cases}$$

$$\exists \lambda_n^2 \in \Lambda \;/\; \begin{cases} \sup_{t\in[0,1]} \left|\lambda_n^2(t) - t\right| \leq \eta, & (2\text{-}i) \\ \sup_{t\in[0,1]} \left|N_{x_n^2}(t) - N_{x_0^2}\left(\lambda_n^2(t)\right)\right| \leq \eta. & (2\text{-}ii) \end{cases}$$

In particular, as η is chosen strictly smaller than 1 and as the $N_{x_n^j}$'s $(n \geq 0,\ j = 1, 2)$ are counting processes with values in $\mathbb{N}$, $(1\text{-}ii)$ implies that $\forall t \in [0,1]$, $N_{x_n^1}(t) = N_{x_0^1}\left(\lambda_n^1(t)\right)$ and thus,

$$u_0 \in x_0^1 \quad \Leftrightarrow \quad u_n = \lambda_n^1(u_0) \in x_n^1.$$

Similarly, (2-*ii*) implies that

$$v_0 \in x_0^2 \quad \Leftrightarrow \quad v_n = \lambda_n^2(v_0) \in x_n^2.$$

Therefore,

$$\begin{aligned}\varphi^w(x_n) &= \iint_{[0,1]^2} w(u,v) dN_{x_n^1}(u) dN_{x_n^2}(v) \\ &= \sum_{(u_n,v_n)\in x_n^1\times x_n^2} w(u_n,v_n) \\ &= \sum_{(u_0,v_0)\in x_0^1\times x_0^2} w\left(\lambda_n^1(u_0),\lambda_n^2(v_0)\right).\end{aligned}$$

Hence,

$$|\varphi^w(x_n) - \varphi^w(x_0)| \leq \sum_{(u_0,v_0)\in x_0^1\times x_0^2} \left|w\left(\lambda_n^1(u_0),\lambda_n^2(v_0)\right) - w\left(u_0,v_0\right)\right|.$$

Yet, by (1-*i*) and (2-*i*), for each (u_0, v_0) in $x_0^1 \times x_0^2$, we have

$$\| \left(\lambda_n^1(u_0),\lambda_n^2(v_0)\right) - (u_0,v_0) \|_\infty \leq \eta,$$

and thus, applying (1.7.1), we obtain

$$|\varphi^w(x_n) - \varphi^w(x_0)| \leq \#x_0^1 \#x_0^2 \varepsilon,$$

and this for all $n \geq n_0$, which ends the proof of Proposition 1.3.1.

1.7.3 Proof of Proposition 1.3.2

Let us prove that in the *Coincidence case*, the kernel $h = h_{\varphi_\delta^{coinc}}$ given by (1.2.1) and (1.2.5) is continuous for the topology induced by the metric d (defined in (1.3.4)) in any (x_0, y_0) in $\mathcal{C}_\delta$ satisfying

$$\left(\left\{x_0^1\right\} \cup \left\{y_0^1\right\}\right) \cap \left(\left\{x_0^2 \pm \delta\right\} \cup \left\{y_0^2 \pm \delta\right\}\right) = \emptyset.$$

As in the proof of Proposition 1.3.1, denote by N_z the counting process associated with z:

$$N_z(t) = \int_0^1 \mathbb{1}_{u\leq t} dN_z(u).$$

Consider a sequence $\{(x_n, y_n)\}_{n\in\mathbb{N}}$ of elements in $\mathcal{X}^2 \times \mathcal{X}^2$, where $x_n = \left(x_n^1, x_n^2\right)$ and $y_n = \left(y_n^1, y_n^2\right)$ such that $d\left((x_n, y_n), (x_0, y_0)\right) \underset{n\to+\infty}{\longrightarrow} 0$ and (x_0, y_0) belongs to $\mathcal{C}_\delta$.
We want to show that $|h(x_n, y_n) - h(x_0, y_0)| \underset{n\to+\infty}{\longrightarrow} 0$.
Since (x_0, y_0) is in $\mathcal{C}_\delta$, for any t_0 in $\left\{x_0^2 \pm \delta\right\} \cup \left\{y_0^2 \pm \delta\right\}$, $t_0 \notin x_0^1$, which means that $N_{x_0^1}$ is continuous in t_0 and therefore constant in a neighborhood:

$$\exists \eta_{t_0} > 0 \ / \ \forall t \in [0,1], \quad |t - t_0| \leq \eta_{t_0} \Rightarrow N_{x_0^1}(t) = N_{x_0^1}(t_0).$$

As $\left\{x_0^2 \pm \delta\right\} \cup \left\{y_0^2 \pm \delta\right\}$ is finite, $\eta_{x_0^1} = \min_{t_0 \in \left\{x_0^2 \pm \delta\right\} \cup \left\{y_0^2 \pm \delta\right\}} \eta_{t_0} > 0$ is well defined, and satisfies

$$\forall u \in \left\{x_0^2 \pm \delta\right\} \cup \left\{y_0^2 \pm \delta\right\}, \ \forall t \in [0,1], \quad |t - u| \leq \eta_{x_0^1} \Rightarrow N_{x_0^1}(t) = N_{x_0^1}(u).$$

By the same argument using continuity of $N_{y_0^1}$ over $\left\{x_0^2 \pm \delta\right\} \cup \left\{y_0^2 \pm \delta\right\}$, one can find $\eta_{y_0^1} > 0$ such that

$$\forall u \in \left\{x_0^2 \pm \delta\right\} \cup \left\{y_0^2 \pm \delta\right\},\ \forall t \in [0,1], \quad |t-u| \leq \eta_{y_0^1} \Rightarrow N_{y_0^1}(t) = N_{y_0^1}(u).$$

Since $(x_0, y_0) \in \mathcal{C}_\delta \Leftrightarrow \left(\left\{x_0^2\right\} \cup \left\{y_0^2\right\}\right) \cap \left(\left\{x_0^1 \pm \delta\right\} \cup \left\{y_0^1 \pm \delta\right\}\right) = \emptyset$, one can construct $\eta_{x_0^2}$ and $\eta_{y_0^2}$ satisfying

$$\forall u \in \left\{x_0^1 \pm \delta\right\} \cup \left\{y_0^1 \pm \delta\right\},\ \forall t \in [0,1], \begin{cases} |t-u| \leq \eta_{x_0^2} \Rightarrow N_{x_0^2}(t) = N_{x_0^2}(u), \\ |t-u| \leq \eta_{y_0^2} \Rightarrow N_{y_0^2}(t) = N_{y_0^2}(u). \end{cases}$$

Finally, if $\eta = \min\left\{\eta_{x_0^1}, \eta_{y_0^1}, \eta_{x_0^2}, \eta_{y_0^2}\right\} > 0$,

$$\forall s \in \left\{x_0^2 \pm \delta\right\} \cup \left\{y_0^2 \pm \delta\right\}, \forall t \in [0,1], |t-s| \leq \eta \Rightarrow \begin{cases} N_{x_0^1}(t) = N_{x_0^1}(s), \\ N_{y_0^1}(t) = N_{y_0^1}(s), \end{cases} \tag{1.7.2}$$

$$\forall s \in \left\{x_0^1 \pm \delta\right\} \cup \left\{y_0^1 \pm \delta\right\}, \forall t \in [0,1], |t-s| \leq \eta \Rightarrow \begin{cases} N_{x_0^2}(t) = N_{x_0^2}(s), \\ N_{y_0^2}(t) = N_{y_0^2}(s). \end{cases} \tag{1.7.3}$$

As $d\left((x_n, y_n), (x_0, y_0)\right) \to_{n \to +\infty} 0$, there exists $n_0 \geq 0$ such that for $n \geq n_0$, $d\left((x_n, y_n), (x_0, y_0)\right) \leq \eta/4$. From the definition of d, we deduce that

$$\exists \lambda_n^1 \in \Lambda \ / \ \begin{cases} \sup_{t \in [0,1]} \left|\lambda_n^1(t) - t\right| \leq \frac{\eta}{4}, & (1\text{-}i) \\ \sup_{t \in [0,1]} \left|N_{x_n^1}(t) - N_{x_0^1}\left(\lambda_n^1(t)\right)\right| \leq \frac{\eta}{4}, & (1\text{-}ii) \end{cases}$$

and

$$\exists \lambda_n^2 \in \Lambda \ / \ \begin{cases} \sup_{t \in [0,1]} \left|\lambda_n^2(t) - t\right| \leq \frac{\eta}{4}, & (2\text{-}i) \\ \sup_{t \in [0,1]} \left|N_{x_n^2}(t) - N_{x_0^2}\left(\lambda_n^2(t)\right)\right| \leq \frac{\eta}{4}. & (2\text{-}ii) \end{cases}$$

Notice that similar results occur for y_n and y_0, but there are not detailed here since we do not use them explicitly.
By definition of h,

$$\begin{aligned} & h(x_n, y_n) - h\left(x_0, y_0\right) \\ &= \frac{1}{2} \iint \mathbb{1}_{|u-v| \leq \delta} \left\{dN_{x_n^1} dN_{x_n^2} + dN_{y_n^1} dN_{y_n^2} - dN_{x_n^1} dN_{y_n^2} - dN_{y_n^1} dN_{x_n^2}\right\}(u,v) \\ &\quad - \frac{1}{2} \iint \mathbb{1}_{|u-v| \leq \delta} \left\{dN_{x_0^1} dN_{x_0^2} + dN_{y_0^1} dN_{y_0^2} - dN_{x_0^1} dN_{y_0^2} - dN_{y_0^1} dN_{x_0^2}\right\}(u,v) \\ &= \frac{1}{2} \iint \mathbb{1}_{|u-v| \leq \delta} \Big(dN_{x_n^1}(u) \left(dN_{x_n^2} - dN_{x_0^2}\right)(v) + dN_{y_n^1}(u) \left(dN_{y_n^2} - dN_{y_0^2}\right)(v) \\ &\quad - dN_{x_n^1}(u) \left(dN_{y_n^2} - dN_{y_0^2}\right)(v) - dN_{y_n^1}(u) \left(dN_{x_n^2} - dN_{x_0^2}\right)(v) \\ &\quad + \left(dN_{x_n^1} - dN_{x_0^1}\right)(u)\ dN_{x_0^2}(v) + \left(dN_{y_n^1} - dN_{y_0^1}\right)(u)\ dN_{y_0^2}(v) \\ &\quad - \left(dN_{x_n^1} - dN_{x_0^1}\right)(u)\ dN_{y_0^2}(v) + \left(dN_{y_n^1} - dN_{y_0^1}\right)(u)\ dN_{x_0^2}(v) \Big). \end{aligned} \tag{1.7.4}$$

By symmetry of the problem, we just need to study the terms

$$A_n = \iint \mathbb{1}_{|u-v| \leq \delta} \left(dN_{x_n^1} - dN_{x_0^1}\right)(u)\ dN_{x_0^2}(v),$$

and

$$B_n = \iint \mathbb{1}_{|u-v| \leq \delta} dN_{x_n^1}(u) \left(dN_{x_n^2} - dN_{x_0^2}\right)(v).$$

Study of A_n.

$$\begin{aligned} A_n &= \iint \mathbb{1}_{|u-v|\le\delta}\left(dN_{x_n^1}-dN_{x_0^1}\right)(u)\; dN_{x_0^2}(v) \\ &= \iint \mathbb{1}_{u\le v+\delta}\left(dN_{x_n^1}-dN_{x_0^1}\right)(u)\; dN_{x_0^2}(v) \\ &\quad - \iint \mathbb{1}_{u<v-\delta}\left(dN_{x_n^1}-dN_{x_0^1}\right)(u)\; dN_{x_0^2}(v). \end{aligned}$$

We have that

$$\begin{aligned} &\left|\iint \mathbb{1}_{u\le v+\delta}\left(dN_{x_n^1}-dN_{x_0^1}\right)(u)\; dN_{x_0^2}(v)\right| \\ &\qquad = \left|\int \left(N_{x_n^1}(v+\delta)-N_{x_0^1}(v+\delta)\right) dN_{x_0^2}(v)\right| \\ &\qquad \le \sum_{T\in x_0^2}\left|N_{x_n^1}(T+\delta)-N_{x_0^1}(T+\delta)\right| \\ &\qquad \le \sum_{T\in x_0^2}\left|N_{x_n^1}(T+\delta)-N_{x_0^1}\left(\lambda_n^1(T+\delta)\right)\right| \\ &\qquad + \sum_{T\in x_0^2}\left|N_{x_0^1}\left(\lambda_n^1(T+\delta)\right)-N_{x_0^1}(T+\delta)\right|. \end{aligned}$$

Now, using the notation $N^-_{x_i^1}(t)=\int \mathbb{1}_{u<t} dN_{x_i^1}(u)$,

$$\left|\iint \mathbb{1}_{u<v-\delta}\left(dN_{x_n^1}-dN_{x_0^1}\right)(u)\; dN_{x_0^2}(v)\right| \le \sum_{T\in x_0^2}\left|N^-_{x_n^1}(T-\delta)-N^-_{x_0^1}(T-\delta)\right|.$$

Therefore,

$$\begin{aligned} |A_n| \le \sum_{T\in x_0^2}\Bigg(&\left|N_{x_n^1}(T+\delta)-N_{x_0^1}\left(\lambda_n^1(T+\delta)\right)\right| \\ &+\left|N_{x_0^1}\left(\lambda_n^1(T+\delta)\right)-N_{x_0^1}(T+\delta)\right| \\ &+\left|N^-_{x_n^1}(T-\delta)-N^-_{x_0^1}(T-\delta)\right|\Bigg). \end{aligned} \tag{1.7.5}$$

Let us study individually each term in the sum.
Fix T in x_0^2. By (1-*ii*),

$$\left|N_{x_n^1}(T+\delta)-N_{x_0^1}\left(\lambda_n^1(T+\delta)\right)\right| \le \frac{\eta}{4} \le \varepsilon. \tag{1.7.6}$$

From (1-*i*), one has $|\lambda_n^1(T+\delta)-(T+\delta)| \le \frac{\eta}{2} \le \eta$ which, with (1.7.2), implies

$$\left|N_{x_0^1}\left(\lambda_n^1(T+\delta)\right)-N_{x_0^1}(T+\delta)\right| = 0. \tag{1.7.7}$$

As $N^-_{x_n^1}(T-\delta) = \lim\limits_{\substack{u\to T-\delta\\ u<T-\delta}} N_{x_n^1}(u)$, there exists u_T in $[T-\delta-\eta/4, T-\delta)$ such that

$$\left|N^-_{x_n^1}(T-\delta)-N_{x_n^1}(u_T)\right| \le \varepsilon,$$

so

$$\left|N^-_{x^1_n}(T-\delta) - N^-_{x^1_0}(T-\delta)\right| \le \varepsilon + \left|N_{x^1_n}(u_T) - N_{x^1_0}\left(\lambda^1_n(u_T)\right)\right| + \left|N_{x^1_0}\left(\lambda^1_n(u_T)\right) - N^-_{x^1_0}(T-\delta)\right|. \quad (1.7.8)$$

From (1-*ii*), one has $\left|N_{x^1_n}(u_T) - N_{x^1_0}\left(\lambda^1_n(u_T)\right)\right| \le \eta/4 \le \varepsilon$.
Then, by continuity of $N_{x^1_0}$ at $T-\delta$, first remark that $N^-_{x^1_0}(T-\delta) = N_{x^1_0}(T-\delta)$. Moreover, by (1-*i*) and construction of u_T,

$$\left|\lambda^1_n(u_T) - (T-\delta)\right| \le \left|\lambda^1_n(u_T) - u_T\right| + |u_T - (T-\delta)| \le \frac{\eta}{4} + \frac{\eta}{4} < \eta,$$

hence, using (1.7.2), $\left|N_{x^1_0}\left(\lambda^1_n(u_T)\right) - N^-_{x^1_0}(T-\delta)\right| = 0$. So finally, (1.7.8) gives

$$\left|N^-_{x^1_n}(T-\delta) - N^-_{x^1_0}(T-\delta)\right| \le 2\varepsilon. \quad (1.7.9)$$

Combining (1.7.5), (1.7.6), (1.7.7), and (1.7.9), we obtain that for any $n \ge n_0$:

$$|A_n| \le 3\varepsilon \# x^2_0. \quad (1.7.10)$$

Study of B_n. Recall that $B_n = \iint \mathbb{1}_{|u-v|\le\delta} dN_{x^1_n}(u)\left(dN_{x^2_n} - dN_{x^2_0}\right)(v)$.
As for A_n, B_n is upper bounded by a sum of several terms, that we study separately.

$$B_n = \sum_{T\in x^1_n}\left(N_{x^2_n}(T+\delta) - N_{x^2_0}(T+\delta)\right) - \sum_{T\in x^1_n}\left(N^-_{x^2_n}(T-\delta) - N^-_{x^2_0}(T-\delta)\right).$$

So

$$B_n \le |B_{n,1}| + |B_{n,2}| + |B_{n,3}| + |B_{n,4}|, \quad (1.7.11)$$

with

$$\begin{aligned}
B_{n,1} &= \sum_{T\in x^1_n}\left(N_{x^2_n}(T+\delta) - N_{x^2_0}\left(\lambda^2_n(T+\delta)\right)\right),\\
B_{n,2} &= \sum_{T\in x^1_n}\left(N_{x^2_0}\left(\lambda^2_n(T+\delta)\right) - N_{x^2_0}(T+\delta)\right),\\
B_{n,3} &= \sum_{T\in x^1_n}\left(N^-_{x^2_n}(T-\delta) - N^-_{x^2_0}\left(\lambda^2_n(T-\delta)\right)\right),\\
B_{n,4} &= \sum_{T\in x^1_n}\left(N^-_{x^2_0}\left(\lambda^2_n(T-\delta)\right) + N^-_{x^2_0}(T-\delta)\right).
\end{aligned}$$

The control of B_n is quite similar to the one of A_n except that the sums are over T in x^1_n instead of T in x^1_0, which prevents us to use (1.7.3) and (1.7.2) directly.

Control of $B_{n,1}$. Due to (2-*ii*), $\left|N_{x^2_n}(T+\delta) - N_{x^2_0}\left(\lambda^2_n(T+\delta)\right)\right| \le \varepsilon$, so

$$|B_{n,1}| \le \varepsilon \# x^1_n. \quad (1.7.12)$$

Control of $B_{n,2}$. One can easily see that

$$\begin{aligned} B_{n,2} &= \iint \left(\mathbb{1}_{v\leq\lambda_n^2(u+\delta)} - \mathbb{1}_{v\leq u+\delta}\right) dN_{x_0^2}(v)\; dN_{x_n^1}(u) \\ &= \iint \left[\left(1-\mathbb{1}_{u<(\lambda_n^2)^{-1}(v)-\delta}\right) - (1-\mathbb{1}_{u<v-\delta})\right] dN_{x_0^2}(v)\; dN_{x_n^1}(u) \\ &= \sum_{T\in x_0^2} \left(N_{x_n^1}^-(T-\delta) - N_{x_n^1}^-\left(\left(\lambda_n^2\right)^{-1}(T)-\delta\right)\right). \end{aligned}$$

Fix now T in x_0^2.

$$\begin{aligned} \left|N_{x_n^1}^-(T-\delta) - N_{x_n^1}^-\left(\left(\lambda_n^2\right)^{-1}(T)-\delta\right)\right| \leq{}& \left|N_{x_n^1}^-(T-\delta) - N_{x_0^1}(T-\delta)\right| \\ &+ \left|N_{x_0^1}(T-\delta) - N_{x_n^1}^-\left(\left(\lambda_n^2\right)^{-1}(T)-\delta\right)\right|. \end{aligned}$$

As shown in (1.7.8), $\left|N_{x_n^1}^-(T-\delta) - N_{x_0^1}(T-\delta)\right| \leq 2\varepsilon$.
Furthermore, take v_T in $\left[\left(\lambda_n^2\right)^{-1}(T)-\delta-\eta/4, \left(\lambda_n^2\right)^{-1}(T)-\delta\right)$ such that

$$\left|N_{x_n^1}^-\left(\left(\lambda_n^2\right)^{-1}(T)-\delta\right) - N_{x_n^1}(v_T)\right| \leq \varepsilon.$$

So,

$$\begin{aligned} \left|N_{x_0^1}(T-\delta) - N_{x_n^1}^-\left(\left(\lambda_n^2\right)^{-1}(T)-\delta\right)\right| \leq{}& \varepsilon + \left|N_{x_n^1}(v_T) - N_{x_0^1}\left(\lambda_n^1(v_T)\right)\right| \\ &+ \left|N_{x_0^1}\left(\lambda_n^1(v_T)\right) - N_{x_0^1}(T-\delta)\right|. \end{aligned}$$

By construction of v_T and λ_n^1 (see (1-*ii*)), $\left|N_{x_n^1}(v_T) - N_{x_0^1}\left(\lambda_n^1(v_T)\right)\right| \leq \varepsilon$.
Because of (1.7.2) which is true as

$$\left|\lambda_n^1(v_T) - (T-\delta)\right| \leq |\lambda_n^1(v_T) - v_T| + |v_T - (T-\delta)| \leq \frac{\eta}{4} + \frac{\eta}{4} < \eta$$

by (1-*i*), $\left|N_{x_0^1}\left(\lambda_n^1(v_T)\right) - N_{x_0^1}(T-\delta)\right| = 0$. Hence,

$$\left|N_{x_0^1}(T-\delta) - N_{x_n^1}^-\left(\left(\lambda_n^2\right)^{-1}(T)-\delta\right)\right| \leq 2\varepsilon.$$

Finally,

$$\left|N_{x_n^1}^-(T-\delta) - N_{x_n^1}^-\left(\left(\lambda_n^2\right)^{-1}(T)-\delta\right)\right| \leq 4\varepsilon,$$

and

$$|B_{n,2}| \leq 4\varepsilon \# x_0^2. \tag{1.7.13}$$

Control of $B_{n,3}$. First, for all T in x_n^1, we find some $\nu_{n,T}$ in $(0, \eta/4]$ such that

$$\forall u \in [T-\delta-\nu_{n,T}, T-\delta), \quad \left|N_{x_n^2}^-(T-\delta) - N_{x_n^2}(u)\right| \leq \varepsilon.$$

Setting $\nu_n = \min_{T\in x_n^1} \nu_{n,T}$,

$$\begin{aligned}|B_{n,3}| \quad \leq \quad & \sum_{T\in x_n^1} \left|N^-_{x_n^2}\left(T-\delta\right) - N_{x_n^2}\left(T-\delta-\nu_n\right)\right| \\ & + \sum_{T\in x_n^1} \left|N_{x_n^2}\left(T-\delta-\nu_n\right) - N_{x_0^2}\left(\lambda_n^2\left(T-\delta-\nu_n\right)\right)\right| \\ & + \left|\sum_{T\in x_n^1} \left(N_{x_0^2}\left(\lambda_n^2\left(T-\delta-\nu_n\right)\right) - N^-_{x_0^2}\left(\lambda_n^2\left(T-\delta\right)\right)\right)\right|.\end{aligned}$$

For each T in x_n^1, $\left|N^-_{x_n^2}\left(T-\delta\right) - N_{x_n^2}\left(T-\delta-\nu_n\right)\right| \leq \varepsilon$ and $\left|N_{x_n^2}\left(T-\delta-\nu_n\right) - N_{x_0^2}\left(\lambda_n^2\left(T-\delta-\nu_n\right)\right)\right| \leq \varepsilon$ by (2-ii). Therefore,

$$|B_{n,3}| \leq 2\varepsilon \# x_n^1 + \left|\sum_{T\in x_n^1} \left(N_{x_0^2}\left(\lambda_n^2\left(T-\delta-\nu_n\right)\right) - N^-_{x_0^2}\left(\lambda_n^2\left(T-\delta\right)\right)\right)\right|.$$

Now,

$$\begin{aligned}& \sum_{T\in x_n^1} \left(N_{x_0^2}\left(\lambda_n^2\left(T-\delta-\nu_n\right)\right) - N^-_{x_0^2}\left(\lambda_n^2\left(T-\delta\right)\right)\right) \\ & \qquad = \iint \mathbb{1}_{v\leq\lambda_n^2(u-\delta-\nu_n)} - \mathbb{1}_{v<\lambda_n^2(u-\delta)} dN_{X_n^1}(u)\; dN_{X_0^2}(v) \\ & \qquad = \sum_{T\in x_0^2} \left(N_{x_n^1}\left(\left(\lambda_n^2\right)^{-1}(T)+\delta\right) - N^-_{x_n^1}\left(\left(\lambda_n^2\right)^{-1}(T)+\delta+\nu_n\right)\right).\end{aligned}$$

For each T in x_0^2,

$$\begin{aligned}& \left|N_{x_n^1}\left(\left(\lambda_n^2\right)^{-1}(T)+\delta\right) - N^-_{x_n^1}\left(\left(\lambda_n^2\right)^{-1}(T)+\delta+\nu_n\right)\right| \\ & \quad \leq \left|N_{x_n^1}\left(\left(\lambda_n^2\right)^{-1}(T)+\delta\right) - N_{x_0^1}\left(\lambda_n^1\left(\left(\lambda_n^2\right)^{-1}(T)+\delta\right)\right)\right| \\ & \qquad + \left|N_{x_0^1}\left(\lambda_n^1\left(\left(\lambda_n^2\right)^{-1}(T)+\delta\right)\right) - N_{x_0^1}\left(\lambda_n^1\left(\left(\lambda_n^2\right)^{-1}(T)+\delta+\nu_n\right)\right)\right| \\ & \qquad + \left|N_{x_0^1}\left(\lambda_n^1\left(\left(\lambda_n^2\right)^{-1}(T)+\delta+\nu_n\right)\right) - N^-_{x_n^1}\left(\left(\lambda_n^2\right)^{-1}(T)+\delta+\nu_n\right)\right| \\ & \quad \leq 2\varepsilon + \left|N_{x_0^1}\left(\lambda_n^1\left(\left(\lambda_n^2\right)^{-1}(T)+\delta+\nu_n\right)\right) - N^-_{x_n^1}\left(\left(\lambda_n^2\right)^{-1}(T)+\delta+\nu_n\right)\right|,\end{aligned}$$

where the last line comes from (1-ii), and (1.7.2).
We now find some w_T in $\left[\left(\lambda_n^2\right)^{-1}(T)+\delta+\nu_n-\eta/4\ ,\ \left(\lambda_n^2\right)^{-1}(T)+\delta+\nu_n\right)$ such that

$$\left|N^-_{x_n^1}\left(\left(\lambda_n^2\right)^{-1}(T)+\delta+\nu_n\right) - N_{x_n^1}\left(w_T\right)\right| \leq \varepsilon,$$

so

$$\begin{aligned}& \left|N_{x_0^1}\left(\lambda_n^1\left(\left(\lambda_n^2\right)^{-1}(T)+\delta+\nu_n\right)\right) - N^-_{x_n^1}\left(\left(\lambda_n^2\right)^{-1}(T)+\delta+\nu_n\right)\right| \\ & \qquad \leq \left|N_{x_0^1}\left(\lambda_n^1\left(\left(\lambda_n^2\right)^{-1}(T)+\delta+\nu_n\right)\right) - N_{x_0^1}\left(\lambda_n^1\left(w_T\right)\right)\right| \\ & \qquad\qquad + \left|N_{x_0^1}\left(\lambda_n^1\left(w_T\right)\right) - N_{x_n^1}\left(w_T\right)\right| + \varepsilon.\end{aligned}$$

From (1-*ii*), we deduce that $\left|N_{x_0^1}\left(\lambda_n^1\left(w_T\right)\right)-N_{x_n^1}\left(w_T\right)\right|\leq\varepsilon$. Due to (1.7.2), (1-*i*), and the construction of w_T,

$$\left|\left(\lambda_n^1\left(\left(\lambda_n^2\right)^{-1}\left(T\right)+\delta+\nu_n\right)\right)-\left(T-\delta\right)\right|\leq\frac{3\eta}{4}<\eta,$$

and

$$\left|\left(\lambda_n^1\left(w_T\right)\right)-\left(T-\delta\right)\right|\leq\left|\left(\lambda_n^1\left(w_T\right)-w_T\right)\right|+\left|w_T-\left(T-\delta\right)\right|<\eta.$$

So $\left|N_{x_0^1}\left(\lambda_n^1\left(\left(\lambda_n^2\right)^{-1}\left(T\right)+\delta+\nu_n\right)\right)-N_{x_0^1}\left(\lambda_n^1\left(w_T\right)\right)\right|=0$. As a consequence,

$$\left|N_{x_n^1}\left(\left(\lambda_n^2\right)^{-1}\left(T\right)+\delta\right)-N_{x_n^1}^{-}\left(\left(\lambda_n^2\right)^{-1}\left(T\right)+\delta+\nu_n\right)\right|\leq4\varepsilon,$$

and

$$\left|B_{n,3}\right|\leq2\varepsilon\#x_n^1+4\varepsilon\#x_0^2. \tag{1.7.14}$$

Control of $B_{n,4}$.

$$\begin{aligned}B_{n,4}&=\iint\left(\mathbb{1}_{v<\lambda_n^2(u-\delta)}-\mathbb{1}_{v<u-\delta}\right)dN_{x_0^2}(v)\;dN_{x_n^1}(u)\\&=\sum_{T\in x_0^2}\left(N_{x_n^1}(T+\delta)-N_{x_n^1}\left(\left(\lambda_n^2\right)^{-1}\left(T\right)+\delta\right)\right).\end{aligned}$$

Let us fix T in x_0^2. We have

$$\begin{aligned}&\left|N_{x_n^1}(T+\delta)-N_{x_n^1}\left(\left(\lambda_n^2\right)^{-1}\left(T\right)+\delta\right)\right|\\&\qquad\leq\left|N_{x_n^1}(T+\delta)-N_{x_0^1}\left(\lambda_n^1\left(T+\delta\right)\right)\right|\\&\qquad\quad+\left|N_{x_0^1}\left(\lambda_n^1\left(T+\delta\right)\right)-N_{x_0^1}\left(\lambda_n^1\left(\left(\lambda_n^2\right)^{-1}\left(T\right)+\delta\right)\right)\right|\\&\qquad\quad+\left|N_{x_0^1}\left(\lambda_n^1\left(\left(\lambda_n^2\right)^{-1}\left(T\right)+\delta\right)\right)-N_{x_n^1}\left(\left(\lambda_n^2\right)^{-1}\left(T\right)+\delta\right)\right|.\end{aligned}$$

The first and the last terms are upper bounded by ε due to (1-*ii*). Furthermore, since $N_{x_0^1}\left(\lambda_n^1\left(\left(\lambda_n^2\right)^{-1}\left(T\right)+\delta\right)\right)=N_{x_0^1}\left(T+\delta\right)=N_{x_0^1}\left(\lambda_n^1\left(T+\delta\right)\right)$ by applying (1.7.2) and using (1-*i*) and (2-*i*),

$$\left|N_{x_0^1}\left(\lambda_n^1\left(T+\delta\right)\right)-N_{x_0^1}\left(\lambda_n^1\left(\left(\lambda_n^2\right)^{-1}\left(T\right)+\delta\right)\right)\right|=0.$$

So finally,

$$\left|B_{n,4}\right|\leq2\varepsilon\#x_0^2. \tag{1.7.15}$$

Combining (1.7.11), (1.7.12), (1.7.13), (1.7.14), and (1.7.15), we can conclude that

$$\left|B_n\right|\leq3\varepsilon\#x_n^1+10\varepsilon\#x_0^2. \tag{1.7.16}$$

We now just remark that $\left(\#x_n^1\right)_{n\geq n_0}$ is bounded because it converges to $\#x_0^1$. Indeed, since $\#x_n^1=N_{x_n^1}(1)$, $\#x_0^1=N_{x_0^1}(1)$ and for every n, $\lambda_n^1(1)=1$,

$$\begin{aligned}\left|\#x_n^1-\#x_0^1\right|&=\left|N_{x_n^1}(1)-N_{x_0^1}(1)\right|\\&=\left|N_{x_n^1}(1)-N_{x_0^1}\left(\lambda_n^1(1)\right)\right|\\&\underset{n\to+\infty}{\longrightarrow}0.\end{aligned}$$

With (1.7.4), (1.7.10), and (1.7.16), this concludes the proof of Proposition 1.3.2.

1.7.4 Proof of Theorem 1.3.1

By Proposition 1.3.3, for all $n \geq 2$,

$$d_2\Big(\mathcal{L}\left(\sqrt{n}U_n, P_n^1 \otimes P_n^2 | \mathbb{X}_n\right), \mathcal{L}\left(\sqrt{n}U_n, P^1 \otimes P^2\right)\Big) \leq C \inf_{\substack{\left(Y_{n,a}^*, Y_a\right), \left(Y_{n,b}^*, Y_b\right) \text{ i.i.d. ;} \\ Y_{n,a}^*, Y_{n,b}^* \sim P_n^1 \otimes P_n^2,\ Y_a, Y_b \sim P^1 \otimes P^2}} \mathbb{E}^*\left[\left(h\left(Y_{n,a}^*, Y_{n,b}^*\right) - h\left(Y_a, Y_b\right)\right)^2\right].$$

Our goal is to construct, for almost all ω in Ω, a sequence of random variables $\left(\bar{Y}_{n,\omega,a}^*\right)_{n\geq 1}$ such that for every $n \geq 1$, $\bar{Y}_{n,\omega,a}^* \sim P_{n,\omega}^1 \otimes P_{n,\omega}^2$, where $P_{n,\omega}^j = n^{-1}\sum_{i=1}^n \delta_{X_i^j(\omega)}$ is the jth marginal empirical measure corresponding to the realization $\mathbb{X}_n(\omega)$, a random variable $\bar{Y}_{\omega,a} \sim P^1 \otimes P^2$, and $\left\{\left(\bar{Y}_{n,\omega,b}^*\right)_{n\geq 1}, \bar{Y}_{\omega,b}\right\}$ an independent copy of $\left\{\left(\bar{Y}_{n,\omega,a}^*\right)_{n\geq 1}, \bar{Y}_{\omega,a}\right\}$ on some probability space $(\Omega'_\omega, \mathcal{A}'_\omega, \mathbb{P}'_\omega)$ depending on ω such that

$$\mathbb{E}'_\omega\left[\left(h\left(\bar{Y}_{n,\omega,a}^*, \bar{Y}_{n,\omega,b}^*\right) - h\left(\bar{Y}_{\omega,a}, \bar{Y}_{\omega,b}\right)\right)^2\right] \underset{n\to+\infty}{\longrightarrow} 0, \tag{1.7.17}$$

where $\mathbb{E}'_\omega$ denotes the expectation corresponding to $\mathbb{P}'_\omega$. Then from (1.7.17), we can conclude by noting that, for almost all ω in Ω,

$$\inf_{\substack{\left(Y_{n,a}^*, Y_a\right), \left(Y_{n,b}^*, Y_b\right) \text{ i.i.d. ;} \\ Y_{n,a}^*, Y_{n,b}^* \sim P_{n,\omega}^1 \otimes P_{n,\omega}^2,\ Y_a, Y_b \sim P^1 \otimes P^2}} \mathbb{E}^*\left[\left(h\left(Y_{n,a}^*, Y_{n,b}^*\right) - h\left(Y_a, Y_b\right)\right)^2\right](\omega) \leq \mathbb{E}'_\omega\left[\left(h\left(\bar{Y}_{n,\omega,a}^*, \bar{Y}_{n,\omega,b}^*\right) - h\left(\bar{Y}_{\omega,a}, \bar{Y}_{\omega,b}\right)\right)^2\right] \underset{n\to+\infty}{\longrightarrow} 0.$$

To prove (1.7.17), consider $(\Omega, \mathcal{A}, \mathbb{P})$ the probability space on which all the X_i's are defined. In what follows, one can keep in mind that Ω represents the randomness in the original sequence $(X_i)_i$. Thus, a given ω in Ω represents a given realization of $(X_i)_i$.
As a preliminary step, from Proposition 1.3.4, there exists some subset Ω_1 of Ω such that $\mathbb{P}(\Omega_1) = 1$ and for every ω in Ω_1,

$$\frac{1}{n^4}\sum_{i,j,k,l=1}^n h^2\left(\left(X_i^1(\omega), X_j^2(\omega)\right), \left(X_k^1(\omega), X_l^2(\omega)\right)\right) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}\left[h^2\left(\left(X_1^1, X_2^2\right), \left(X_3^1, X_4^2\right)\right)\right]. \tag{1.7.18}$$

Applying Theorem 3 in [174] (see Appendix A.1.2, Theorem A.1.4), since $(\mathcal{X}, d_\mathcal{X})$ defined by (1.3.3) is separable, P-a.s. in $(X_i)_i$, $P_n^1 \underset{n\to+\infty}{\Longrightarrow} P^1$ and $P_n^2 \underset{n\to+\infty}{\Longrightarrow} P^2$. Hence there exists some subset Ω_2 of Ω such that $\mathbb{P}(\Omega_2) = 1$ and for every ω in Ω_2,

$$P_{n,\omega}^1 \otimes P_{n,\omega}^2 \underset{n\to+\infty}{\Longrightarrow} P^1 \otimes P^2, \tag{1.7.19}$$

Now, consider $\Omega_0 = \Omega_1 \cap \Omega_2$, and fix ω in Ω_0.
Following the proof of Skorokhod's representation theorem in [46, Theorem 11.7.2, p. 415] (see Appendix A.1, Theorem A.1.2), since $(\mathcal{X}^2, d_{\mathcal{X}^2})$ is a separable space, it is possible to construct

- some probability space $(\Omega'_\omega, \mathcal{A}'_\omega, \mathbb{P}'_\omega)$,
- some random variables $\bar{Y}^*_{n,\omega,a} : \Omega'_\omega \to \mathcal{X}^2$, $\bar{Y}^*_{n,\omega,b} : \Omega'_\omega \to \mathcal{X}^2$ with distribution $P^1_{n,\omega} \otimes P^2_{n,\omega}$,
- $\bar{Y}_{\omega,a} : \Omega'_\omega \to \mathcal{X}^2$, $\bar{Y}_{\omega,b} : \Omega'_\omega \to \mathcal{X}^2$ with distribution $P^1 \otimes P^2$,

satisfying:

- $\mathbb{P}'_\omega$-a.s., $\bar{Y}^*_{n,\omega,a} \underset{n\to+\infty}{\longrightarrow} \bar{Y}_{\omega,a}$ and $\bar{Y}^*_{n,\omega,b} \underset{n\to+\infty}{\longrightarrow} \bar{Y}_{\omega,b}$,
- $\left\{ \left(\bar{Y}^*_{n,\omega,a}\right)_{n\geq 1}, \bar{Y}_{\omega,a} \right\}$ and $\left\{ \left(\bar{Y}^*_{n,\omega,b}\right)_{n\geq 1}, \bar{Y}_{\omega,b} \right\}$ are independent,

so that w.r.t. the metric d (see (1.3.4)),

$$\mathbb{P}'_\omega\text{-a.s.}, \quad \left(\bar{Y}^*_{n,\omega,a}, \bar{Y}^*_{n,\omega,b}\right) \underset{n\to+\infty}{\longrightarrow} \left(\bar{Y}_{\omega,a}, \bar{Y}_{\omega,b}\right). \tag{1.7.20}$$

But under condition $(\mathcal{A}_{Cont})$, h is continuous on a subset $\mathcal{C}$ of $\mathcal{X}^2$ such that $\mathbb{P}'_\omega\left(\left(\bar{Y}_{\omega,a}, \bar{Y}_{\omega,b}\right) \in \mathcal{C}\right) = \left(P^1 \otimes P^2\right)^{\otimes 2}(\mathcal{C}) = 1$, hence

$$\mathbb{P}'_\omega\text{-a.s.}, \; h\left(\bar{Y}^*_{n,\omega,a}, \bar{Y}^*_{n,\omega,b}\right) \to_{n\to+\infty} h\left(\bar{Y}_{\omega,a}, \bar{Y}_{\omega,b}\right).$$

As $\mathbb{P}'_\omega$-a.s. convergence implies convergence in probability, to obtain (1.7.17), we only need to prove that the sequence $\left(h^2\left(\bar{Y}^*_{n,\omega,a}, \bar{Y}^*_{n,\omega,b}\right)\right)_{n\geq 1}$ is uniformly integrable, according to Theorem 16.6 of [162, p. 165]. We therefore conclude since (1.7.18) is equivalent to

$$\begin{aligned}\mathbb{E}'_\omega\left[h^2\left(\bar{Y}^*_{n,\omega,a}, \bar{Y}^*_{n,\omega,b}\right)\right] &= \frac{1}{n^4}\sum_{i,j,k,l=1}^{n} h^2\left((X^1_i(\omega), X^2_j(\omega)), (X^1_k(\omega), X^2_l(\omega))\right) \\ &\underset{n\to+\infty}{\longrightarrow} \mathbb{E}\left[h^2\left(\left(X^1_1, X^2_2\right), \left(X^1_3, X^2_4\right)\right)\right] = \mathbb{E}'_\omega\left[h^2\left(\bar{Y}_{\omega,a}, \bar{Y}_{\omega,b}\right)\right].\end{aligned}$$

(1.7.17) is thus obtained for any ω in Ω_0, with $\mathbb{P}(\Omega_0) = 1$. This ends the proof.

1.7.5 Proof of Proposition 1.3.3

Fix some integer $n \geq 2$ and recall that the P^j_n's $(j = 1, 2)$ are the marginal empirical measures associated with $\mathbb{X}_n$.
Let $(Y^*_{n,i}, Y_i)_{1\leq i\leq n}$ be an i.i.d. sample such that for every $i = 1 \ldots n$, $Y^*_{n,i} \sim P^1_n \otimes P^2_n$, $Y_i \sim P^1 \otimes P^2$, and such that, from the definition of Wasserstein's metric d_2 recalled in (1.3.2),

$$\begin{aligned}&d_2^2\left(\mathcal{L}\left(\sqrt{n}U_n, P^1_n \otimes P^2_n \middle| \mathbb{X}_n\right), \mathcal{L}\left(\sqrt{n}U_n, P^1 \otimes P^2\right)\right) \\ &\qquad \leq \frac{1}{n(n-1)^2}\,\mathbb{E}^*\left[\left(\sum_{i\neq i'}\left(h\left(Y^*_{n,i}, Y^*_{n,i'}\right) - h\left(Y_i, Y_{i'}\right)\right)\right)^2\right].\end{aligned}$$

Notice that the upper-bound is finite under $(\mathcal{A}^*_{Mmt})$.
Introducing for (i, i', j, j') in $\{1, 2, \ldots, n\}^4$, and m in $\{2, 3, 4\}$,

$$\mathrm{E}_{(i,i',j,j')} = \mathbb{E}^*\Big[\Big(h\big(Y^*_{n,i}, Y^*_{n,i'}\big) - h(Y_i, Y_{i'})\Big)\Big(h\big(Y^*_{n,j}, Y^*_{n,j'}\big) - h(Y_j, Y_{j'})\Big)\Big],$$

$$I_m = \left\{(i,i',j,j') \in \{1,2,\ldots,n\}^4 ; i \neq i',\ j \neq j',\ \#\{i,i',j,j'\} = m\right\},$$

where $\#\{i,i',j,j'\}$ denotes the number of different elements in $\{i,i',j,j'\}$, one has

$$\mathbb{E}^*\left[\left(\sum_{i\neq i'}\left(h\left(Y^*_{n,i},Y^*_{n,i'}\right)-h\left(Y_i,Y_{i'}\right)\right)\right)^2\right] = \sum_{(i,i',j,j')\in I_4} \mathrm{E}_{(i,i',j,j')} + \sum_{(i,i',j,j')\in I_3} \mathrm{E}_{(i,i',j,j')} + \sum_{(i,i',j,j')\in I_2} \mathrm{E}_{(i,i',j,j')}.$$

Let us now upper bound each term of this sum separately.
If (i,i',j,j') is in I_4, then by independence,

$$\mathrm{E}_{(i,i',j,j')} = \left(\mathbb{E}^*\left[h\left(Y^*_{n,i},Y^*_{n,i'}\right)\right] - \mathbb{E}\left[h\left(Y_i,Y_{i'}\right)\right]\right) \times \left(\mathbb{E}^*\left[h\left(Y^*_{n,j},Y^*_{n,j'}\right)\right] - \mathbb{E}\left[h\left(Y_j,Y_{j'}\right)\right]\right).$$

Under $(\mathcal{A}_{Cent})$ and $(\mathcal{A}^*_{Cent})$, $\mathbb{E}\left[h(Y_i,Y_{i'})\right] = \mathbb{E}^*\left[h\left(Y^*_{n,i},Y^*_{n,i'}\right)\right] = 0$, so $\mathrm{E}_{(i,i',j,j')} = 0$.
If (i,i',j,j') is in I_3, by the Cauchy-Schwarz inequality,

$$\mathrm{E}_{(i,i',j,j')} \leq \mathbb{E}^*\left[\left(h\left(Y^*_{n,a},Y^*_{n,b}\right)-h(Y_a,Y_b)\right)^2\right],$$

where $\left(Y^*_{n,a},Y_a\right)$ and $\left(Y^*_{n,b},Y_b\right)$ are independent copies of the $(Y_{n,i},Y_i)$'s. If (i,i',j,j') is in I_2, then $\mathrm{E}_{(i,i',j,j')} = \mathbb{E}^*\left[\left(h\left(Y^*_{n,a},Y^*_{n,b}\right)-h\left(Y_a,Y_b\right)\right)^2\right]$ is immediate.
But $\#I_3 = 4n(n-1)(n-2)$ and $\#I_2 = 2n(n-1)$, so

$$d_2^2\left(\mathcal{L}\left(\sqrt{n}U_n,P_n^1\otimes P_n^2\middle|\,\mathbb{X}_n\right),\mathcal{L}\left(\sqrt{n}U_n,P^1\otimes P^2\right)\right) \leq 4\mathbb{E}^*\left[\left(h\left(Y^*_{n,a},Y^*_{n,b}\right)-h\left(Y_a,Y_b\right)\right)^2\right].$$

Since $\left(Y^*_{n,a},Y_a\right)$ and $\left(Y^*_{n,b},Y_b\right)$ may be arbitrarily chosen, Proposition 1.3.3 follows.

1.7.6 Proof of Proposition 1.3.4

Let us first notice that (1.3.6) is a direct application of the strong law of large numbers for U-statistics, proved by Hoeffding [85] (see Appendix A.3.2, Theorem A.3.1).
Next, for m in $\{1,\ldots,4\}$, introduce

$$g_m\left(X_{i_1},\ldots,X_{i_m}\right) = \sum_{(i,j,k,l)\in I_{\{i_1,\ldots i_m\}}} h^2\left(\left(X_i^1,X_j^2\right),\left(X_k^1,X_l^2\right)\right),$$

where $I_{\{i_1,\ldots i_m\}}$ is the set $\left\{(i,j,k,l)\in\{i_1,\ldots i_m\}^4 ; \#\{i,j,k,l\} = m\right\}$.
Then,

$$\frac{1}{n^4}\sum_{i,j,k,l=1}^{n} h^2\left(\left(X_i^1,X_j^2\right),\left(X_k^1,X_l^2\right)\right) = \sum_{m=1}^{4}\frac{1}{m!}\left(\frac{1}{n^4}\sum_{\substack{(i_1,\ldots,i_m)\in\{1,\ldots,n\}^m\\ i_1,\ldots,i_m \text{ all different}}} g_m\left(X_{i_1},\ldots,X_{i_m}\right)\right).$$

Each of the four terms in the right-hand side of the above decomposition being, up to a multiplicative factor, a classical U-statistic, and since under $(\mathcal{A}^*_{Mmt})$, one has

$\mathbb{E}\left[|g_m(X_{i_1},\dots,X_{i_m})|\right] < +\infty$, we can now apply the strong law of large numbers for U-statistics again. Therefore P-a.s. in $(X_i)_i$,

$$\frac{1}{n(n-1)\dots(n-m+1)}\sum_{(i_1,\dots,i_m)} g_m\left(X_{i_1},\dots,X_{i_m}\right) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}\left[g_m\left(X_1,\dots,X_m\right)\right].$$

In particular, P-a.s. in $(X_i)_i$, $n^{-4}\sum_{(i_1,\dots,i_m)} g_m\left(X_{i_1},\dots,X_{i_m}\right)$ converges towards 0 for m in $\{1,2,3\}$, and towards $\mathbb{E}\left[g_4\left(X_1,X_2,X_3,X_4\right)\right]$ for $m=4$. Finally noticing that

$$\mathbb{E}\left[g_4\left(X_1,X_2,X_3,X_4\right)\right] = 4!\mathbb{E}\left[h^2\left(\left(X_1^1,X_2^2\right),\left(X_3^1,X_4^2\right)\right)\right]$$

allows to conclude.

1.7.7 Proof of Proposition 1.3.5

Let $(X_i)_i$ be a sequence of i.i.d. pairs of point processes with distribution $P^1\otimes P^2$ on $\mathcal{X}^2$. According to $(\mathcal{A}_{Cent})$, for $i\neq j$, $\mathbb{E}\left[h(X_i,X_j)\right]=0$. For a better readability, we set

$$\mathbb{E}\left[h|X_i\right] = \mathbb{E}\left[h(X_i,X)|X_i\right] = \mathbb{E}\left[h(X,X_i)|X_i\right],$$

for some X with distribution $P^1\otimes P^2$, and independent of X_i. By Hoeffding's decomposition for non-degenerate U-statistics (see Appendix A.3.2), which also holds when the X_i's are not necessarily real-valued (see [158]) we obtain that

$$\sqrt{n}U_n(\mathbb{X}_n) = \frac{2}{\sqrt{n}(n-1)}\left(T_n+M_n\right),$$

where

$$T_n = \sum_{i<j}\left(\mathbb{E}\left[h|X_i\right]+\mathbb{E}\left[h|X_j\right]\right), \text{ and } M_n = \sum_{i<j} g(X_i,X_j),$$

with $g(X_i,X_j) = h(X_i,X_j) - \mathbb{E}\left[h|X_i\right] - \mathbb{E}\left[h|X_j\right]$.
Firstly, we have that

$$\mathbb{E}\left[M_n^2\right] = \sum_{i<j}\sum_{k<l}\mathbb{E}\left[g(X_i,X_j)g(X_k,X_l)\right].$$

But if $\{i,j\}\cap\{k,l\}=\emptyset$, $i<j$, $k<l$, then $\mathbb{E}\left[g(X_i,X_j)g(X_k,X_l)\right] = \left(\mathbb{E}\left[g(X_i,X_j)\right]\right)^2 = 0$, and if $\#(\{i,j\}\cap\{k,l\})=1$, with for instance $k=i$, $j\neq l$, $(i<j,\ i<l)$ (the other cases may be treated similarly), then $\mathbb{E}\left[g(X_i,X_j)g(X_i,X_l)\right] = \mathbb{E}\left[\mathbb{E}\left[g(X_i,X_j)|X_i\right]\mathbb{E}\left[g(X_i,X_l)|X_i\right]\right] = 0$. Therefore,

$$\mathbb{E}\left[M_n^2\right] = \sum_{i<j}\mathbb{E}\left[g^2(X_i,X_j)\right] = n(n-1)\mathbb{E}\left[g^2(X_1,X_2)\right]/2,$$

and since $\mathbb{E}\left[g^2(X_i,X_j)\right]<+\infty$, from Chebychev's inequality, we deduce that

$$\frac{2}{\sqrt{n}(n-1)}M_n \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0. \tag{1.7.21}$$

Secondly, we have that $T_n = (n-1)\sum_{i=1}^n \mathbb{E}\left[h|X_i\right]$. Since the $\mathbb{E}\left[h|X_i\right]$'s are i.i.d., with $\mathbb{E}\left[\mathbb{E}\left[h|X_i\right]\right]=0$ and $\mathrm{Var}(\mathbb{E}\left[h|X_i\right]) = \sigma^2_{P^1\otimes P^2}/4$, thanks to $(\mathcal{A}_{Mmt})$, the central limit theorem leads to

$$\frac{2}{\sqrt{n}(n-1)}T_n \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\left(0,\sigma^2_{P^1\otimes P^2}\right). \tag{1.7.22}$$

Thus, combining (1.7.21) and (1.7.22), Slutsky's lemma (see Appendix A.1, Proposition A.1.1) ensures the convergence in distribution of $\sqrt{n}U_n(\mathbb{X}_n)$ towards $\mathcal{N}(0, \sigma^2_{P^1\otimes P^2})$.
Now, in order to obtain the convergence in the Wasserstein metric, one needs to check the convergence of the second-order moments (see Appendix A.1.2, Proposition A.1.2). Notice that

$$\mathbb{E}\left[\left(\sqrt{n}U_n(\mathbb{X}_n)\right)^2\right] = \frac{1}{n(n-1)^2}\sum_{i\neq i'}\sum_{j\neq j'}\mathbb{E}\left[h(X_i,X_{i'})h(X_j,X_{j'})\right].$$

Let us consider all the cases where $i \neq i'$ and $j \neq j'$.
If $\#\{i,i',j,j'\} = 4$, $\mathbb{E}\left[h(X_i,X_{i'})h(X_j,X_{j'})\right] = 0$, by independence and $(\mathcal{A}_{Cent})$.
If $\#\{i,i',j,j'\} = 3$, $\mathbb{E}\left[h(X_i,X_{i'})h(X_j,X_{j'})\right] = \sigma^2_{P^1\otimes P^2}/4$, by symmetry of h.
If $\#\{i,i',j,j'\} = 2$, $\mathbb{E}\left[h(X_i,X_{i'})h(X_j,X_{j'})\right] = \mathbb{E}\left[\left(h(X_1,X_2)\right)^2\right]$. Therefore,

$$\mathbb{E}\left[\left(\sqrt{n}U_n(\mathbb{X}_n)\right)^2\right] = \frac{n-2}{n-1}\sigma^2_{P^1\otimes P^2} + \frac{2}{n-1}\mathbb{E}\left[\left(h(X_1,X_2)\right)^2\right] \underset{n\to+\infty}{\longrightarrow} \sigma^2_{P^1\otimes P^2},$$

which ends the proof of Proposition 1.3.5.

1.7.8 Proof of Corollary 1.3.1

By Proposition 1.3.5, we have that

$$\mathcal{L}\left(\sqrt{n}U_n, P^1\otimes P^2\right) \underset{n\to+\infty}{\Longrightarrow} \mathcal{N}(0,\sigma^2_{P^1\otimes P^2}), \tag{1.7.23}$$

where $\mathcal{N}(0,\sigma^2_{P^1\otimes P^2})$ has a continuous c.d.f. Therefore, by [172, Lemma 2.11] (see Appendix A.1, Lemma A.1.1),

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}\left(\sqrt{n}U_n(\mathbb{X}_n^{\perp\!\!\!\perp}) \leq z\right) - \Phi_{0,\sigma^2_{P^1\otimes P^2}}(z)\right| \underset{n\to+\infty}{\longrightarrow} 0. \tag{1.7.24}$$

Furthermore, since convergence w.r.t. the d_2 distance implies weak convergence, Theorem 1.3.1 combined with (1.7.23) leads to

$$\mathcal{L}\left(\sqrt{n}U_n, P_n^1\otimes P_n^2 \middle|\, \mathbb{X}_n\right) \underset{n\to+\infty}{\Longrightarrow} \mathcal{N}(0,\sigma^2_{P^1\otimes P^2}) \quad P\text{-a.s. in } (X_i)_i. \tag{1.7.25}$$

Hence,

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}\left(\sqrt{n}U_n(\mathbb{X}_n^*)\leq z|\mathbb{X}_n\right) - \Phi_{0,\sigma^2_{P^1\otimes P^2}}(z)\right| \underset{n\to+\infty}{\longrightarrow} 0 \ P\text{-a.s. in } (X_i)_i, \tag{1.7.26}$$

and the first part of the corollary is obtained.

Moreover, [172, Lemma 21.2] (see Appendix A.1, Lemma A.1.2) can then be applied to both (1.7.23) and (1.7.25), to obtain that on the event where (1.7.25) holds :

$$q^*_{\eta,n}(\mathbb{X}_n) \underset{n\to+\infty}{\longrightarrow} \Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(\eta) \quad P\text{-a.s. in } (X_i)_i, \tag{1.7.27}$$

and that $q^{\perp\!\!\!\perp}_{\eta,n}$ also converges to $\Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(\eta)$.

1.7.9 Proof of Theorem 1.3.2

Let us focus on the sequence of upper-tailed tests in $\Gamma(q^*)$, the proof for the other tests being similar.
Under $(\mathcal{H}_0)$, from Proposition 1.3.5 and (1.7.27), by Slutsky's lemma (Appendix A.1, Proposition A.1.1), $(\sqrt{n}U_n(\mathbb{X}_n), q^*_{1-\alpha,n}(\mathbb{X}_n))$ converges in distribution to $(Z, \Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(1-\alpha))$, where $Z \sim \mathcal{N}(0, \sigma^2_{P^1\otimes P^2})$. Therefore, under $(\mathcal{H}_0)$,

$$\mathbb{P}(\sqrt{n}U_n(\mathbb{X}_n) > q^*_{1-\alpha,n}(\mathbb{X}_n)) \to_{n\to+\infty} \alpha,$$

which proves $(\mathcal{P}_{size})$.
Under any alternative such that $\int h(x,x')dP(x)dP(x') > 0$, by Proposition 1.3.4,

$$U_n(\mathbb{X}_n) \underset{n\to+\infty}{\longrightarrow} \int h(x,x')dP(x)dP(x') > 0, \ P\text{-a.s. in } (X_i)_i.$$

Furthermore, due to (1.7.27), $q^*_{1-\alpha,n}(\mathbb{X}_n)/\sqrt{n} \to_{n\to+\infty} 0$ P-a.s. in $(X_i)_i$.
Hence, $\mathbb{P}(\sqrt{n}U_n(\mathbb{X}_n) \le q^*_{1-\alpha,n}(\mathbb{X}_n)) \to_{n\to+\infty} 0$, and thus $(\mathcal{P}_{consist.})$ is proved.

1.7.10 Proof of Proposition 1.3.6

As above, we focus on the sequence of upper-tailed tests in $\Gamma(q^*_{MC})$. Let $Z \sim \mathcal{N}(0,1)$ and define for z in $\mathbb{R}$,

$$F^*_{n,\mathbb{X}_n}(z) = \mathbb{P}\left(\sqrt{n}U_n(\mathbb{X}^*_n) \le z | \mathbb{X}_n\right), \ F^{*B_n}_{n,\mathbb{X}_n}(z) = \frac{1}{B_n}\sum_{b=1}^{B_n} \mathbb{1}_{\sqrt{n}U_n(\mathbb{X}^{*b}_n)\le z}.$$

By the Dvoretzky-Kiefer-Wolfowitz inequality (see [122, Corollary 1], or Appendix A.1.2, Theorem A.1.5), for $n \ge 2$ and $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}\left(\sup_{z\in\mathbb{R}}\left|F^{*B_n}_{n,\mathbb{X}_n}(z) - F^*_{n,\mathbb{X}_n}(z)\right| > \varepsilon\right) &= \mathbb{E}\left[\mathbb{P}\left(\sup_{z\in\mathbb{R}}\left|F^{*,B_n}_{n,\mathbb{X}_n}(z) - F^*_{n,\mathbb{X}_n}(z)\right| > \varepsilon \middle| \mathbb{X}_n\right)\right] \\ &\le 2e^{-2B_n\varepsilon^2} \underset{n\to+\infty}{\longrightarrow} 0,\end{aligned}$$

that is $\sup_{z\in\mathbb{R}}|F^{*B_n}_{n,\mathbb{X}_n}(z) - F^*_{n,\mathbb{X}_n}(z)| \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0$. With (1.7.26), this leads to

$$\sup_{z\in\mathbb{R}}\left|F^{*B_n}_{n,\mathbb{X}_n}(z) - \Phi_{0,\sigma^2_{P^1\otimes P^2}}(z)\right| \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0. \tag{1.7.28}$$

We finish the proof using similar arguments as in [172, Lemma 21.2] (see Appendix A.1, Lemma A.1.2), combined with a subsequence argument [46, Theorem 9.2.1] (recalled in Appendix A.1, Theorem A.1.1). Let ϕ_0 be an extraction. Then, by (1.7.28), there exists an extraction ϕ_1, and some $\Omega_0 \subset \Omega$ such that $\mathbb{P}(\Omega_0) = 1$, and for every ω in Ω_0,

$$\sup_{z\in\mathbb{R}}\left|F^{*B_{\phi_1\circ\phi_0(n)}}_{\phi_1\circ\phi_0(n),\mathbb{X}_{\phi_1\circ\phi_0(n)}}(\omega)(z) - \Phi_{0,\sigma^2_{P^1\otimes P^2}}(z)\right| \underset{n\to+\infty}{\longrightarrow} 0.$$

From now on, fix ω in Ω_0. In particular, this fixes a realisation of $\mathbb{X}_n$, and a realisation of $\left(\mathbb{X}^{*1}_n, \ldots, \mathbb{X}^{*B_n}_n\right)$ and thus, $F^{*B_n}_{n,\mathbb{X}_n}(\omega)$ is deterministic.

Hence, $F^{*B_{\phi_1\circ\phi_0(n)}}_{\phi_1\circ\phi_0(n),\mathbb{X}_{\phi_1\circ\phi_0(n)}}(\omega)(Z) \xrightarrow[n\to+\infty]{a.s.} \Phi_{0,\sigma^2_{P^1\otimes P^2}}(Z)$, and for η in $(0,1)$,

$$\Phi_{0,1}\left(\left(F^{*B_{\phi_1\circ\phi_0(n)}}_{\phi_1\circ\phi_0(n),\mathbb{X}_{\phi_1\circ\phi_0(n)}}(\omega)\right)^{-1}(\eta)\right) = \mathbb{P}\left(F^{*B_{\phi_1\circ\phi_0(n)}}_{\phi_1\circ\phi_0(n),\mathbb{X}_{\phi_1\circ\phi_0(n)}}(\omega)(Z) < \eta\right)$$
$$\xrightarrow[n\to+\infty]{} \mathbb{P}\left(\Phi_{0,\sigma^2_{P^1\otimes P^2}}(Z) < \eta\right) = \Phi_{0,1}\left(\left(\Phi_{0,\sigma^2_{P^1\otimes P^2}}\right)^{-1}(\eta)\right).$$

Finally, as $\Phi_{0,1}$ is a one-to-one function and $\Phi^{-1}_{0,1}$ is continuous,

$$\sqrt{\phi_1\circ\phi_0(n)}U^{*(\lceil\eta(B_{\phi_1\circ\phi_0(n)})\rceil)}(\omega)$$
$$= \left(F^{*B_{\phi_1\circ\phi_0(n)}}_{\phi_1\circ\phi_0(n),\mathbb{X}_{\phi_1\circ\phi_0(n)}}(\omega)\right)^{-1}(\eta) \xrightarrow[n\to+\infty]{} \Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(\eta), \quad (1.7.29)$$

and this for all ω in Ω_0, and any initial extraction ϕ_0. Therefore, we obtain that

$$\sqrt{n}U^{*(\lceil\eta B_n\rceil)} \xrightarrow[n\to+\infty]{\mathbb{P}} \Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(\eta),$$

and we conclude as for Theorem 1.3.2.

1.7.11 Proof of Theorem 1.4.1

For the sake of clarity and a better readability, we first present a sketch of the proof of this Theorem in Section 1.7.11. A complete version is detailed in Section 1.7.11.

Sketch of proof of Theorem 1.4.1

Let d_{BL} denote the bounded Lipschitz metric, which metrizes the weak convergence (see [46, Prop. 11.3.2 and Th. 11.3.3] or Appendix A.1.2 for more details). For any variable Z_n depending on $\mathbb{X}_n$ and Π_n, $\mathcal{L}(Z_n|\mathbb{X}_n)$ denotes the conditional distribution of Z_n given $\mathbb{X}_n$ and for any integrable function f, $\mathbb{E}_{P^1\otimes P^2}[f] = \mathbb{E}[f(X^1_1, X^2_2)]$.

$\diamond$ The first step of the proof consists in decomposing $\sqrt{n}U_n\left(\mathbb{X}^{\Pi_n}_n\right)$ in

$$\sqrt{n}U_n\left(\mathbb{X}^{\Pi_n}_n\right) = \frac{n}{n-1}\left(M^{\Pi_n}_n(\mathbb{X}_n) + \frac{R^{\Pi_n}_n(\mathbb{X}_n)}{\sqrt{n}} - \frac{T_n(\mathbb{X}_n)}{\sqrt{n}}\right),$$

where

- $M^{\Pi_n}_n(\mathbb{X}_n) = \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i\neq j}\mathbb{1}_{\Pi_n(i)=j}C_{i,j}$,
- $R^{\Pi_n}_n(\mathbb{X}_n) = \displaystyle\sum_{i=1}^{n}\left(\mathbb{1}_{\Pi_n(i)=i} - \frac{1}{n}\right)C_{i,i}$,
- $T_n(\mathbb{X}_n) = \dfrac{1}{n}\displaystyle\sum_{i\neq j}C_{i,j}$,

with

$$C_{i,j} = \varphi\left(X^1_i, X^2_j\right) - \mathbb{E}\left[\varphi\left(X^1_i, X^2\right)\middle|X^1_i\right] - \mathbb{E}\left[\varphi\left(X^1, X^2_j\right)\middle|X^2_j\right] + \mathbb{E}_{P^1\otimes P^2}[\varphi],$$

$X = (X^1, X^2)$ being P-distributed and independent of $(X_i)_i$.
We then prove from Cauchy-Schwarz inequality that

$$\mathbb{E}\left[\left(\mathbb{E}\left[\frac{\left|R_n^{\Pi_n}(\mathbb{X}_n)\right|}{\sqrt{n}}\middle|\mathbb{X}_n\right]\right)^2\right] \underset{n\to+\infty}{\longrightarrow} 0 \text{ and } \mathbb{E}\left[\left(\frac{T_n(\mathbb{X}_n)}{\sqrt{n}}\right)^2\right] \underset{n\to+\infty}{\longrightarrow} 0,$$

therefore from Markov's inequality,

$$\mathbb{E}\left[\frac{\left|R_n^{\Pi_n}(\mathbb{X}_n)\right|}{\sqrt{n}}\middle|\mathbb{X}_n\right] \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0 \text{ and } \frac{T_n(\mathbb{X}_n)}{\sqrt{n}} \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0.$$

From the definition of d_{BL}, this allows us to derive that

$$d_{BL}\left(\mathcal{L}\left(\sqrt{n}U_n\left(\mathbb{X}_n^{\Pi_n}\right)\middle|\mathbb{X}_n\right), \mathcal{L}\left(\frac{n}{n-1}M_n^{\Pi_n}(\mathbb{X}_n)\middle|\mathbb{X}_n\right)\right) \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0. \tag{1.7.30}$$

$\diamond$ The second, and most difficult, step of the proof consists in proving that

$$d_{BL}\left(\mathcal{L}\left(M_n^{\Pi_n}(\mathbb{X}_n)\middle|\mathbb{X}_n\right), \mathcal{N}\left(0, \sigma^2_{P^1\otimes P^2}\right)\right) \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0. \tag{1.7.31}$$

Consider

$$Y_{n,i} = \frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}\left(\mathbb{1}_{\Pi_n(i)=j}C_{i,j} + \mathbb{1}_{\Pi_n(j)=i}C_{j,i}\right), \tag{1.7.32}$$

and for Π'_n another uniformly distributed random permutation with values in $\mathfrak{S}_n$, independent of Π_n and $\mathbb{X}_n$, define accordingly $Y'_{n,i}$ by replacing Π_n by Π'_n in (1.7.32), so that $M_n^{\Pi_n}(\mathbb{X}_n) = \sum_{i=1}^n Y_{n,i}$ and similarly for $M_n^{\Pi'_n}(\mathbb{X}_n)$.
Setting $\mathcal{F}_{n,i} = \sigma(\Pi_n, \Pi'_n, X_1, X_2, \ldots, X_i)$ for $n \geq i \geq 2$, we prove through technical computations that for a, b in $\mathbb{R}$, $\left(aY_{n,i} + bY'_{n,i}, \mathcal{F}_{n,i}\right)_{2\leq i\leq n}$ is a martingale difference array which satisfies the assumptions of the following result, commonly attributed to Brown [31].

Theorem 1.7.1. *Let $(X_{n,k})_{k\in\{1,\ldots,p_n\},n\in\mathbb{N}^*}$ be a martingale difference array, that is such that there exists an array of σ-algebra $(\mathcal{F}_{n,k})_{k\in\{1,\ldots,p_n\},n\in\mathbb{N}^*}$ that is increasing w.r.t. k such that for all $k = 1, ..., p_n$, $\mathbb{E}[X_{n,k}|\mathcal{F}_{n,k-1}] = 0$.*
Let $A_n = \sum_{k=1}^{p_n}\mathbb{E}\left[X_{n,k}^2|\mathcal{F}_{n,k-1}\right]$, and assume that

- $A_n \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} \sigma^2 > 0$,
- $\forall\varepsilon > 0$, $\sum_{k=1}^{p_n}\mathbb{E}\left[X_{n,k}^2\mathbb{1}_{|X_{n,k}|>\varepsilon}\right] \underset{n\to+\infty}{\longrightarrow} 0$.

Then $Z_n = \sum_{k=1}^{p_n} X_{n,k}$ converges in distribution towards $\mathcal{N}(0, \sigma^2)$.

Thus, given a, b in $\mathbb{R}$, we obtain that

$$\mathcal{L}\left(aM_n^{\Pi_n}(\mathbb{X}_n) + bM_n^{\Pi'_n}(\mathbb{X}_n)\right) \underset{n\to+\infty}{\Longrightarrow} \mathcal{N}\left(0, \left(a^2+b^2\right)\sigma^2_{P^1\otimes P^2}\right),$$

which, according to the Cramér-Wold device (see Appendix A.1.2, Theorem A.1.3), leads to Lemma 1.7.1 below.

Lemma 1.7.1. *Considering the above notation,*

$$\mathcal{L}\left(\left(M_n^{\Pi_n}\left(\mathbb{X}_n\right), M_n^{\Pi'_n}\left(\mathbb{X}_n\right)\right)'\right) \underset{n\to+\infty}{\Longrightarrow} \mathcal{N}_2\left(0, \begin{pmatrix} \sigma^2_{P^1\otimes P^2} & 0 \\ 0 & \sigma^2_{P^1\otimes P^2} \end{pmatrix}\right),$$

where $\mathcal{N}_2\left(M, V\right)$ *denotes the* 2*-dimensional Gaussian distribution with mean vector* M *and variance-covariance matrix* V*.*

From Lemma 1.7.1, we deduce that for every t in $\mathbb{R}$,

$$\begin{cases} \mathbb{P}\left(M_n^{\Pi_n}\left(\mathbb{X}_n\right) \leq t\right) \underset{n\to+\infty}{\longrightarrow} \Phi_{0,\sigma^2_{P^1\otimes P^2}}(t), \\ \mathbb{P}\left(M_n^{\Pi_n}\left(\mathbb{X}_n\right) \leq t, M_n^{\Pi'_n}\left(\mathbb{X}_n\right) \leq t\right) \underset{n\to+\infty}{\longrightarrow} \Phi^2_{0,\sigma^2_{P^1\otimes P^2}}(t). \end{cases}$$

Using Chebychev's inequality, with the fact that in a separable metric space, convergence in probability is metrizable, and therefore is equivalent to almost sure convergence of a subsequence of any initial subsequence (see, e.g., [46, Th. 9.2.1] recalled in Appendix A.1, Theorem A.1.1), we prove that this leads to (1.7.31), and therefore,

$$d_{BL}\left(\mathcal{L}\left(\sqrt{n}U_n\left(\mathbb{X}_n^{\Pi_n}\right)\middle|\mathbb{X}_n\right), \mathcal{N}\left(0, \sigma^2_{P^1\otimes P^2}\right)\right) \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0.$$

$\diamond$ The third, and final, step of the proof consists in deriving, by direct computations and the strong law of large numbers of Hœffding [85] (see Appendix A.3.2, Theorem A.3.1), the convergence of the conditional second-order moments

$$\mathbb{E}\left[\left(\sqrt{n}U_n\left(\mathbb{X}_n^{\Pi_n}\right)\right)^2\middle|\mathbb{X}_n\right] \underset{n\to+\infty}{\overset{a.s.}{\longrightarrow}} \sigma^2_{P^1\otimes P^2}, \tag{1.7.33}$$

which ends the proof.

Complete proof of Theorem 1.4.1

Recall that d_{BL} denotes the bounded Lipschitz metric which metrizes the weak convergence, defined by

$$d_{BL}(\mu,\nu) = \sup_{f\in BL,\ \|f\|_{BL}\leq 1}\left|\int_{\mathbb{R}} f\ (d\mu - d\nu)\right|,$$

where, as defined in [46] or in Appendix A.1.2, BL is the set of bounded Lipschitz functions on $\mathbb{R}$, and

$$\|f\|_{BL} = \|f\|_\infty + \sup_{x\neq y}\frac{|f(x)-f(y)|}{|x-y|}.$$

Recall that the proof consists of three steps presented in Section 1.7.11. We give below a complete proof for each of these steps.

First step: decomposition of $\sqrt{n}U_n\left(\mathbb{X}_n^{\Pi_n}\right)$ in the *Linear case*. It is obvious that by the definition (1.2.6) of h_φ,

$$U_n\left(\mathbb{X}_n^{\Pi_n}\right) = \frac{1}{n-1}U_n^{\Pi_n}, \tag{1.7.34}$$

where $U_n^{\Pi_n} = \sum_{i=1}^n \varphi\Big(X_i^1, X_{\Pi_n(i)}^2\Big) - \frac{1}{n}\sum_{i,j=1}^n \varphi(X_i^1, X_j^2)$, so,

$$\begin{aligned} U_n^{\Pi_n} &= \sum_{i=1}^n \varphi\Big(X_i^1, X_{\Pi_n(i)}^2\Big) - \frac{1}{n}\sum_{i,j=1}^n \mathbb{E}\left[\varphi(X_i^1, X_j^2)\middle|X_i^1\right] \\ &\quad -\frac{1}{n}\sum_{i,j=1}^n \mathbb{E}\left[\varphi(X_i^1, X_j^2)\middle|X_j^2\right] + \frac{1}{n}\sum_{i,j=1}^n \mathbb{E}\left[\varphi(X_i^1, X_j^2)\right] \\ &\quad -\frac{1}{n}\sum_{i,j=1}^n \left(\varphi(X_i^1, X_j^2) - \mathbb{E}\left[\varphi(X_i^1, X_j^2)\middle|X_i^1\right]\right. \\ &\quad \left.-\mathbb{E}\left[\varphi(X_i^1, X_j^2)\middle|X_j^2\right] + \mathbb{E}\left[\varphi(X_i^1, X_j^2)\right]\right). \end{aligned}$$

On the one hand, if $\mathbb{E}_P[f]$ and $\mathbb{E}_{P^1\otimes P^2}[f]$ respectively denote $\mathbb{E}\left[f\left(X_1^1, X_1^2\right)\right]$, and $\mathbb{E}\left[f\left(X_1^1, X_2^2\right)\right]$, for any integrable function f, then

$$\begin{aligned} &\frac{1}{n}\sum_{i,j=1}^n \mathbb{E}\left[\varphi(X_i^1, X_j^2)\right] \\ &= \sum_{i,j=1}^n \mathbb{1}_{\Pi_n(i)=j}\mathbb{E}\left[\varphi(X_i^1, X_j^2)\right] - \sum_{i,j=1}^n \left(\mathbb{1}_{\Pi_n(i)=j} - \frac{1}{n}\right)\mathbb{E}\left[\varphi(X_i^1, X_j^2)\right] \\ &= \sum_{i,j=1}^n \mathbb{1}_{\Pi_n(i)=j}\mathbb{E}\left[\varphi(X_i^1, X_j^2)\right] - (\mathbb{E}_P[\varphi] - \mathbb{E}_{P^1\otimes P^2}[\varphi])\sum_{i=1}^n\left(\mathbb{1}_{\Pi_n(i)=i} - \frac{1}{n}\right). \end{aligned}$$

On the other hand,

$$\begin{aligned} \frac{1}{n}\sum_{i,j=1}^n \mathbb{E}\left[\varphi(X_i^1, X_j^2)\middle|X_i^1\right] &= \sum_{i,j=1}^n \mathbb{1}_{\Pi_n(i)=j}\mathbb{E}\left[\varphi(X_i^1, X_j^2)\middle|X_i^1\right] \\ &\quad - \sum_{i=1}^n \left(\mathbb{1}_{\Pi_n(i)=i} - \frac{1}{n}\right)\left(\mathbb{E}\left[\varphi(X_i^1, X_i^2)\middle|X_i^1\right] - \mathbb{E}\left[\varphi(X_i^1, X^2)\middle|X_i^1\right]\right), \end{aligned}$$

where $X = (X^1, X^2)$ is assumed to be P-distributed and independent of $(X_i)_i$, and in the same way,

$$\begin{aligned} \frac{1}{n}\sum_{i,j=1}^n \mathbb{E}\left[\varphi(X_i^1, X_j^2)\middle|X_j^2\right] &= \sum_{i,j=1}^n \mathbb{1}_{\Pi_n(i)=j}\mathbb{E}\left[\varphi(X_i^1, X_j^2)\middle|X_j^2\right] \\ &\quad - \sum_{j=1}^n \left(\mathbb{1}_{\Pi_n(j)=j} - \frac{1}{n}\right)\left(\mathbb{E}\left[\varphi(X_j^1, X_j^2)\middle|X_j^2\right] - \mathbb{E}\left[\varphi(X^1, X_j^2)\middle|X_j^2\right]\right). \end{aligned}$$

Therefore, $U_n^{\Pi_n}$ is equal to

$$\sum_{i,j=1}^{n} \mathbb{1}_{\Pi_n(i)=j}\Big(\varphi\big(X_i^1,X_j^2\big) - \mathbb{E}\left[\varphi\big(X_i^1,X_j^2\big)\big|X_i^1\right] - \mathbb{E}\left[\varphi\big(X_i^1,X_j^2\big)\big|X_j^2\right] + \mathbb{E}\left[\varphi\big(X_i^1,X_j^2\big)\right]\Big)$$
$$+ \sum_{i=1}^{n}\left(\mathbb{1}_{\Pi_n(i)=i} - \frac{1}{n}\right)\Big(\mathbb{E}\left[\varphi\big(X_i^1,X_i^2\big)\big|X_i^1\right] + \mathbb{E}\left[\varphi\big(X_i^1,X_i^2\big)\big|X_i^2\right]$$
$$- \mathbb{E}\left[\varphi\big(X_i^1,X^2\big)\big|X_i^1\right] - \mathbb{E}\left[\varphi\big(X^1,X_i^2\big)\big|X_i^2\right] - \mathbb{E}_P\left[\varphi\right] + \mathbb{E}_{P^1\otimes P^2}\left[\varphi\right]\Big)$$
$$- \frac{1}{n}\sum_{i,j=1}^{n}\Big(\varphi\big(X_i^1,X_j^2\big) - \mathbb{E}\left[\varphi\big(X_i^1,X_j^2\big)\big|X_i^1\right] - \mathbb{E}\left[\varphi\big(X_i^1,X_j^2\big)\big|X_j^2\right] + \mathbb{E}\left[\varphi\big(X_i^1,X_j^2\big)\right]\Big).$$

As a consequence, setting

$$C_{i,j} = \varphi\big(X_i^1,X_j^2\big) - \mathbb{E}\left[\varphi\big(X_i^1,X^2\big)\big|X_i^1\right] - \mathbb{E}\left[\varphi\big(X^1,X_j^2\big)\big|X_j^2\right] + \mathbb{E}_{P^1\otimes P^2}\left[\varphi\right],$$

$$\sqrt{n}U_n\big(\mathbb{X}_n^{\Pi_n}\big) = \frac{n}{n-1}\left(M_n^{\Pi_n}(\mathbb{X}_n) + \frac{R_n^{\Pi_n}(\mathbb{X}_n)}{\sqrt{n}} - \frac{T_n(\mathbb{X}_n)}{\sqrt{n}}\right), \tag{1.7.35}$$

with

$$\begin{aligned} M_n^{\Pi_n}(\mathbb{X}_n) &= \frac{1}{\sqrt{n}}\sum_{i\neq j}\mathbb{1}_{\Pi_n(i)=j}C_{i,j},\\ R_n^{\Pi_n}(\mathbb{X}_n) &= \sum_{i=1}^{n}\left(\mathbb{1}_{\Pi_n(i)=i} - \frac{1}{n}\right)C_{i,i},\\ T_n(\mathbb{X}_n) &= \frac{1}{n}\sum_{i\neq j}C_{i,j}. \end{aligned}$$

Let us now prove that

$$d_{BL}\left(\mathcal{L}\left(\sqrt{n}U_n\big(\mathbb{X}_n^{\Pi_n}\big)\big|\mathbb{X}_n\right),\mathcal{L}\left(\frac{n}{n-1}M_n^{\Pi_n}(\mathbb{X}_n)\Big|\mathbb{X}_n\right)\right) \xrightarrow[n\to+\infty]{\mathbb{P}} 0. \tag{1.7.36}$$

To do this, first notice that for every function f in BL such that $\|f\|_{BL}\leq 1$,

$$\begin{aligned} &\left|\mathbb{E}\left[f\left(\sqrt{n}U_n\big(\mathbb{X}_n^{\Pi_n}\big)\right)\big|\mathbb{X}_n\right] - \mathbb{E}\left[f\left(\frac{n}{n-1}M_n^{\Pi_n}(\mathbb{X}_n)\right)\Big|\mathbb{X}_n\right]\right|\\ &\qquad\leq \mathbb{E}\left[\left|\sqrt{n}U_n\big(\mathbb{X}_n^{\Pi_n}\big) - \frac{n}{n-1}M_n^{\Pi_n}(\mathbb{X}_n)\right|\Big|\mathbb{X}_n\right]\\ &\qquad\leq \frac{n}{n-1}\left(\mathbb{E}\left[\frac{\left|R_n^{\Pi_n}(\mathbb{X}_n)\right|}{\sqrt{n}}\Big|\mathbb{X}_n\right] + \frac{|T_n(\mathbb{X}_n)|}{\sqrt{n}}\right). \end{aligned}$$

Hence, taking the supremum over $\{f\in BL;\|f\|_{BL}\leq 1\}$,

$$\begin{aligned} &d_{BL}\left(\mathcal{L}\left(\sqrt{n}U_n\big(\mathbb{X}_n^{\Pi_n}\big)\big|\mathbb{X}_n\right),\mathcal{L}\left(\frac{n}{n-1}M_n^{\Pi_n}(\mathbb{X}_n)\Big|\mathbb{X}_n\right)\right)\\ &\qquad\leq \frac{n}{n-1}\left(\mathbb{E}\left[\frac{\left|R_n^{\Pi_n}(\mathbb{X}_n)\right|}{\sqrt{n}}\Big|\mathbb{X}_n\right] + \frac{|T_n(\mathbb{X}_n)|}{\sqrt{n}}\right). \end{aligned} \tag{1.7.37}$$

Moreover, on the one hand, since Π_n is independent of $(X_i)_i$, by Cauchy-Schwarz inequality,

$$\mathbb{E}\left[\left(\mathbb{E}\left[\frac{\left|R_n^{\Pi_n}(\mathbb{X}_n)\right|}{\sqrt{n}}\middle|\mathbb{X}_n\right]\right)^2\right] \leq \frac{1}{n}\mathbb{E}\left[\left(R_n^{\Pi_n}(\mathbb{X}_n)\right)^2\right],$$

and

$$\begin{aligned}
\mathbb{E}\left[\left(R_n^{\Pi_n}(\mathbb{X}_n)\right)^2\right] &\leq \sum_{i,j=1}^{n}\mathbb{E}\left[\left(\mathbb{1}_{\Pi_n(i)=i}-\frac{1}{n}\right)\left(\mathbb{1}_{\Pi_n(j)=j}-\frac{1}{n}\right)\right]\mathbb{E}\left[C_{i,i}C_{j,j}\right] \\
&\leq C\left(\mathbb{E}_P\left[\varphi^2\right]+\mathbb{E}_{P^1\otimes P^2}\left[\varphi^2\right]\right)\sum_{i,j=1}^{n}\left(\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=i}\mathbb{1}_{\Pi_n(j)=j}\right]-\frac{1}{n^2}\right) \\
&\leq C\left(\mathbb{E}_P\left[\varphi^2\right]+\mathbb{E}_{P^1\otimes P^2}\left[\varphi^2\right]\right)\left(\sum_{i=1}^{n}\left(\frac{1}{n}-\frac{1}{n^2}\right)+\sum_{i\neq j}\left(\frac{1}{n(n-1)}-\frac{1}{n^2}\right)\right) \\
&\leq C\left(\mathbb{E}_P\left[\varphi^2\right]+\mathbb{E}_{P^1\otimes P^2}\left[\varphi^2\right]\right) < +\infty.
\end{aligned}$$

Therefore, from Markov's inequality, we deduce that

$$\mathbb{E}\left[\frac{\left|R_n^{\Pi_n}(\mathbb{X}_n)\right|}{\sqrt{n}}\middle|\mathbb{X}_n\right] \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

On the other hand,

$$\mathbb{E}\left[\left(\frac{T_n(\mathbb{X}_n)}{\sqrt{n}}\right)^2\right] = \frac{1}{n^3}\sum_{i\neq j}\sum_{k\neq l}\mathbb{E}\left[C_{i,j}C_{k,l}\right].$$

Notice that for $i\neq j$, $\mathbb{E}\left[C_{i,j}|X_i\right]=\mathbb{E}\left[C_{i,j}|X_j\right]=0$.
If $\#\{i,j,k,l\}=4$, then $\mathbb{E}\left[C_{i,j}C_{k,l}\right]=\left(\mathbb{E}\left[C_{i,j}\right]\right)^2=0$.
If i,j,l are all different, then

$$\begin{aligned}
\mathbb{E}\left[C_{i,j}C_{i,l}\right] &= \mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{i,l}|X_i,X_l\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[C_{i,j}|X_i\right]C_{i,l}\right] \\
&= 0.
\end{aligned}$$

In the same way, for i,j,k all different, then $\mathbb{E}\left[C_{i,j}C_{k,i}\right]=0$.
If $i\neq j$,

$$\mathbb{E}\left[C_{i,j}^2\right]=\sigma^2_{P^1\otimes P^2}, \text{ and } \mathbb{E}\left[C_{i,j}C_{j,i}\right]\leq\sigma^2_{P^1\otimes P^2}, \tag{1.7.38}$$

by the Cauchy-Schwarz inequality. Combining the above computations, we obtain that

$$\mathbb{E}\left[\left(\frac{T_n(\mathbb{X}_n)}{\sqrt{n}}\right)^2\right] \leq 2\frac{n(n-1)}{n^3}\sigma^2_{P^1\otimes P^2} \underset{n\to+\infty}{\longrightarrow} 0,$$

and therefore,

$$\frac{T_n(\mathbb{X}_n)}{\sqrt{n}} \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Finally, from (1.7.37), we derive (1.7.36).

Second step: asymptotic normality of $M_n^{\Pi_n}(\mathbb{X}_n)$ given $\mathbb{X}_n$, in probability. Recall that

$$\begin{aligned} M_n^{\Pi_n}(\mathbb{X}_n) &= \frac{1}{\sqrt{n}} \sum_{i \neq j} \mathbb{1}_{\Pi_n(i)=j} C_{i,j} \\ &= \frac{1}{\sqrt{n}} \sum_{i=2}^{n} \sum_{j=1}^{i-1} \left(\mathbb{1}_{\Pi_n(i)=j} C_{i,j} + \mathbb{1}_{\Pi_n(j)=i} C_{j,i} \right). \end{aligned}$$

Let Π_n' be another uniformly distributed random permutation with values in $\mathfrak{S}_n$, independent of Π_n and $\mathbb{X}_n$, and

$$\begin{aligned} M_n^{\Pi_n'}(\mathbb{X}_n) &= \frac{1}{\sqrt{n}} \sum_{i \neq j} \mathbb{1}_{\Pi_n'(i)=j} C_{i,j} \\ &= \frac{1}{\sqrt{n}} \sum_{i=2}^{n} \sum_{j=1}^{i-1} \left(\mathbb{1}_{\Pi_n'(i)=j} C_{i,j} + \mathbb{1}_{\Pi_n'(j)=i} C_{j,i} \right). \end{aligned}$$

Let us now recall the result of Lemma 1.7.1:

$$\mathcal{L}\left(\left(M_n^{\Pi_n}(\mathbb{X}_n), M_n^{\Pi_n'}(\mathbb{X}_n) \right)' \right) \underset{n \to +\infty}{\Longrightarrow} \mathcal{N}_2\left(0, \begin{pmatrix} \sigma^2_{P^1 \otimes P^2} & 0 \\ 0 & \sigma^2_{P^1 \otimes P^2} \end{pmatrix} \right).$$

Proof of Lemma 1.7.1. According to the Cramér-Wold device (see Appendix A.1.2, Theorem A.1.3), given a, b in $\mathbb{R}$, we aim at proving that

$$\mathcal{L}\left(a M_n^{\Pi_n}(\mathbb{X}_n) + b M_n^{\Pi_n'}(\mathbb{X}_n) \right) \underset{n \to +\infty}{\Longrightarrow} \mathcal{N}\left(0, \left(a^2 + b^2 \right) \sigma^2_{P^1 \otimes P^2} \right).$$

In order to deal with simpler mathematical expressions, we introduce below some additional notation.

- For $n \geq i \geq 2$, $\mathcal{F}_{n,i} = \sigma\left(\Pi_n, \Pi_n', X_1, X_2, \ldots, X_i \right)$.
- Let

$$Y_{n,i} = \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} \left(\mathbb{1}_{\Pi_n(i)=j} C_{i,j} + \mathbb{1}_{\Pi_n(j)=i} C_{j,i} \right),$$

$$Y'_{n,i} = \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} \left(\mathbb{1}_{\Pi_n'(i)=j} C_{i,j} + \mathbb{1}_{\Pi_n'(j)=i} C_{j,i} \right),$$

so that $M_n^{\Pi_n}(\mathbb{X}_n) = \sum_{i=1}^n Y_{n,i}$ and $M_n^{\Pi_n'}(\mathbb{X}_n) = \sum_{i=1}^n Y'_{n,i}$.

Let us first prove that for a fixed integer $n \geq 2$, $\left(a Y_{n,i} + b Y'_{n,i}, \mathcal{F}_{n,i} \right)_{2 \leq i \leq n}$ is a martingale difference array. Note that for $2 \leq i \leq n$,

$$\begin{aligned} \mathbb{E}\left[Y_{n,i} | \mathcal{F}_{n,i-1} \right] &= \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} \mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j} C_{i,j} + \mathbb{1}_{\Pi_n(j)=i} C_{j,i} \middle| \mathcal{F}_{n,i-1} \right] \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} \left(\mathbb{1}_{\Pi_n(i)=j} \mathbb{E}\left[C_{i,j} | X_j \right] + \mathbb{1}_{\Pi_n(j)=i} \mathbb{E}\left[C_{j,i} | X_j \right] \right) \\ &= 0. \end{aligned}$$

In the same way, we have that $\mathbb{E}\left[Y'_{n,i}\middle|\mathcal{F}_{n,i-1}\right] = 0$, so $\mathbb{E}\left[aY_{n,i} + bY'_{n,i}\middle|\mathcal{F}_{n,i-1}\right] = 0$. From Theorem 1.7.1, we thus deduce that if

$$(i) \qquad \sum_{i=2}^{n} \mathbb{E}\left[\left(aY_{n,i} + bY'_{n,i}\right)^2\middle|\mathcal{F}_{n,i-1}\right] \xrightarrow[n\to+\infty]{\mathbb{P}} (a^2+b^2)\sigma^2_{P^1\otimes P^2},$$

$$(ii) \qquad \sum_{i=2}^{n} \mathbb{E}\left[\left(aY_{n,i} + bY'_{n,i}\right)^2 \mathbb{1}_{|aY_{n,i}+bY'_{n,i}|>\varepsilon}\right] \xrightarrow[n\to+\infty]{} 0 \text{ for any } \varepsilon > 0,$$

then

$$\mathcal{L}\left(aM_n^{\Pi_n}(\mathbb{X}_n) + bM_n^{\Pi'_n}(\mathbb{X}_n)\right) \underset{n\to+\infty}{\Longrightarrow} \mathcal{N}\left(0, \left(a^2+b^2\right)\sigma^2_{P^1\otimes P^2}\right).$$

Let us now check that both (i) and (ii) are satisfied.

Assumption (i). In all the following, only consider $n \geq 4$. Noticing that

$$\sum_{i=2}^{n} \mathbb{E}\left[\left(aY_{n,i} + bY'_{n,i}\right)^2\middle|\mathcal{F}_{n,i-1}\right] = (a^2+b^2)\sum_{i=2}^{n}\mathbb{E}\left[Y^2_{n,i}\middle|\mathcal{F}_{n,i-1}\right] + 2ab\sum_{i=2}^{n}\mathbb{E}\left[Y_{n,i}Y'_{n,i}\middle|\mathcal{F}_{n,i-1}\right], \quad (1.7.39)$$

the proof of (i) can be decomposed into two points.
The first point consists in proving that

$$\sum_{i=2}^{n}\mathbb{E}\left[Y^2_{n,i}\right] \xrightarrow[n\to+\infty]{} \sigma^2_{P^1\otimes P^2} \quad \text{and} \quad \mathrm{Var}\left(\sum_{i=2}^{n}\mathbb{E}\left[Y^2_{n,i}\middle|\mathcal{F}_{n,i-1}\right]\right) \xrightarrow[n\to+\infty]{} 0,$$

which leads, thanks to Chebychev's inequality, to

$$\sum_{i=2}^{n}\mathbb{E}\left[Y^2_{n,i}\middle|\mathcal{F}_{n,i-1}\right] \xrightarrow[n\to+\infty]{\mathbb{P}} \sigma^2_{P^1\otimes P^2}.$$

The second point consists in proving that

$$\mathbb{E}\left[\left(\sum_{i=2}^{n}\mathbb{E}\left[Y_{n,i}Y'_{n,i}\middle|\mathcal{F}_{n,i-1}\right]\right)^2\right] \xrightarrow[n\to+\infty]{} 0,$$

so

$$\sum_{i=2}^{n}\mathbb{E}\left[Y_{n,i}Y'_{n,i}\middle|\mathcal{F}_{n,i-1}\right] \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

• First point. On the one hand,

$$\sum_{i=2}^{n}\mathbb{E}\left[Y^2_{n,i}\right] = \frac{1}{n}\sum_{i=2}^{n}\sum_{j,k=1}^{i-1}\mathbb{E}\left[\left(\mathbb{1}_{\Pi_n(i)=j}C_{i,j} + \mathbb{1}_{\Pi_n(j)=i}C_{j,i}\right)\left(\mathbb{1}_{\Pi_n(i)=k}C_{i,k} + \mathbb{1}_{\Pi_n(k)=i}C_{k,i}\right)\right].$$

Furthermore, if $1 \leq j \neq k \leq i-1$,

$$\begin{aligned}
&\mathbb{E}\left[\left(\mathbb{1}_{\Pi_n(i)=j}C_{i,j} + \mathbb{1}_{\Pi_n(j)=i}C_{j,i}\right)\left(\mathbb{1}_{\Pi_n(i)=k}C_{i,k} + \mathbb{1}_{\Pi_n(k)=i}C_{k,i}\right)\right] \\
&\quad = \mathbb{E}\left[\mathbb{E}\left[\left(\mathbb{1}_{\Pi_n(i)=j}C_{i,j} + \mathbb{1}_{\Pi_n(j)=i}C_{j,i}\right)\left(\mathbb{1}_{\Pi_n(i)=k}C_{i,k} + \mathbb{1}_{\Pi_n(k)=i}C_{k,i}\right)\middle|X_i, X_j, \Pi_n\right]\right] \\
&\quad = \mathbb{E}\left[\left(\mathbb{1}_{\Pi_n(i)=j}C_{i,j} + \mathbb{1}_{\Pi_n(j)=i}C_{j,i}\right)\left(\mathbb{1}_{\Pi_n(i)=k}\mathbb{E}\left[C_{i,k}|X_i\right] + \mathbb{1}_{\Pi_n(k)=i}\mathbb{E}\left[C_{k,i}|X_i\right]\right)\right] \\
&\quad = 0.
\end{aligned}$$

Thus,

$$\begin{aligned}
\sum_{i=2}^{n}\mathbb{E}\left[Y_{n,i}^2\right] &= \frac{1}{n}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\mathbb{E}\left[\left(\mathbb{1}_{\Pi_n(i)=j}C_{i,j}+\mathbb{1}_{\Pi_n(j)=i}C_{j,i}\right)^2\right]\\
&= \frac{1}{n}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}C_{i,j}^2+\mathbb{1}_{\Pi_n(j)=i}C_{j,i}^2+2\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(j)=i}C_{i,j}C_{j,i}\right]\\
&= \frac{1}{n}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\left(\frac{2}{n}\mathbb{E}\left[C_{i,j}^2\right]+\frac{2}{n(n-1)}\mathbb{E}\left[C_{i,j}C_{j,i}\right]\right)\\
&= \frac{2}{n^2}\sum_{i=2}^{n}(i-1)\left(\mathbb{E}\left[C_{1,2}^2\right]+\frac{1}{n-1}\mathbb{E}\left[C_{1,2}C_{2,1}\right]\right),
\end{aligned}$$

so $\sum_{i=2}^{n}\mathbb{E}\left[Y_{n,i}^2\right]=\frac{n-1}{n}\mathbb{E}\left[C_{1,2}^2\right]+\frac{1}{n}\mathbb{E}\left[C_{1,2}C_{2,1}\right]$. From (1.7.38), we derive that

$$\sum_{i=2}^{n}\mathbb{E}\left[Y_{n,i}^2\right]\underset{n\to+\infty}{\longrightarrow}\sigma_{P^1\otimes P^2}^2. \tag{1.7.40}$$

On the other hand, we have that

$$\begin{aligned}
\mathbb{E}\left[Y_{n,i}^2|\mathcal{F}_{n,i-1}\right] &= \frac{1}{n}\sum_{j=1}^{i-1}\mathbb{1}_{\Pi_n(i)=j}\mathbb{E}\left[C_{i,j}^2|X_j\right]+\frac{1}{n}\sum_{j=1}^{i-1}\mathbb{1}_{\Pi_n(j)=i}\mathbb{E}\left[C_{j,i}^2|X_j\right]\\
&\quad+\frac{2}{n}\sum_{j=1}^{i-1}\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(j)=i}\mathbb{E}\left[C_{i,j}C_{j,i}|X_j\right]\\
&\quad+\frac{2}{n}\sum_{1\le j\ne k\le i-1}\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=i}\mathbb{E}\left[C_{i,j}C_{k,i}|X_j,X_k\right].
\end{aligned}$$

Then,

$$\sum_{i=2}^{n}\left(\mathbb{E}\left[Y_{n,i}^2|\mathcal{F}_{n,i-1}\right]-\mathbb{E}\left[Y_{n,i}^2\right]\right)=A_{n,1}+A_{n,2}+2A_{n,3}+2A_{n,4},$$

with

$$\begin{aligned}
A_{n,1} &= \frac{1}{n}\sum_{1\le j<i\le n}\left(\mathbb{1}_{\Pi_n(i)=j}\mathbb{E}\left[C_{i,j}^2|X_j\right]-\frac{1}{n}\mathbb{E}\left[C_{i,j}^2\right]\right),\\
A_{n,2} &= \frac{1}{n}\sum_{1\le j<i\le n}\left(\mathbb{1}_{\Pi_n(j)=i}\mathbb{E}\left[C_{j,i}^2|X_j\right]-\frac{1}{n}\mathbb{E}\left[C_{i,j}^2\right]\right),\\
A_{n,3} &= \frac{1}{n}\sum_{1\le j<i\le n}\left(\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(j)=i}\mathbb{E}\left[C_{i,j}C_{j,i}|X_j\right]-\frac{1}{n(n-1)}\mathbb{E}\left[C_{i,j}C_{j,i}\right]\right),\\
A_{n,4} &= \frac{1}{n}\sum_{1\le j\ne k<i\le n}\left(\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=i}\mathbb{E}\left[C_{i,j}C_{k,i}|X_j,X_k\right]\right).
\end{aligned}$$

Thus,

$$\mathrm{Var}\left(\sum_{i=2}^{n}\left(\mathbb{E}\left[Y_{n,i}^2|\mathcal{F}_{n,i-1}\right]\right)\right)\le 4\left(\mathbb{E}\left[A_{n,1}^2\right]+\mathbb{E}\left[A_{n,2}^2\right]+4\mathbb{E}\left[A_{n,3}^2\right]+4\mathbb{E}\left[A_{n,4}^2\right]\right). \tag{1.7.41}$$

Let us now control each term of the above right-hand side.
Convergence of $\mathbb{E}\left[A_{n,1}^2\right]$ *and* $\mathbb{E}\left[A_{n,2}^2\right]$.

$$\mathbb{E}\left[A_{n,1}^2\right] = \frac{1}{n^2}\sum_{1\le j<i\le n}\sum_{1\le l<k\le n}\Big(\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=l}\right]\times \mathbb{E}\left[\mathbb{E}\left[C_{i,j}^2|X_j\right]\mathbb{E}\left[C_{k,l}^2|X_l\right]\right] - \frac{1}{n^2}\left(\mathbb{E}\left[C_{k,l}^2\right]\right)^2\Big).$$

Let us now consider all the cases where $1\le j<i\le n$, and $1\le l<k\le n$.
If $i=k$ and $j=l$, then

$$\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=l}\right]\mathbb{E}\left[\mathbb{E}\left[C_{i,j}^2|X_j\right]\mathbb{E}\left[C_{k,l}^2|X_l\right]\right] = \frac{1}{n}\mathbb{E}\left[\left(\mathbb{E}\left[C_{2,1}^2|X_1\right]\right)^2\right].$$

If $i=k$ and $j\neq l$, or if $i\neq k$ and $j=l$, then

$$\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=l}\right]\mathbb{E}\left[\mathbb{E}\left[C_{i,j}^2|X_j\right]\mathbb{E}\left[C_{k,l}^2|X_l\right]\right] = 0.$$

If $i\neq k$ and $j\neq l$, then

$$\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=l}\right]\mathbb{E}\left[\mathbb{E}\left[C_{i,j}^2|X_j\right]\mathbb{E}\left[C_{k,l}^2|X_l\right]\right] = \frac{1}{n(n-1)}\left(\mathbb{E}\left[C_{2,1}^2\right]\right)^2.$$

By combining these results, from (1.7.38) and under the assumption $(\mathcal{A}_{\varphi,Mmt})$, we obtain that

$$\mathbb{E}\left[A_{n,1}^2\right] \le \frac{n-1}{2n^2}\left(\mathbb{E}\left[\left(\mathbb{E}\left[C_{2,1}^2|X_1\right]\right)^2\right] - \frac{\sigma_{P^1\otimes P^2}^4}{n}\right) + C\,n^2\left(\frac{1}{n(n-1)}-\frac{1}{n^2}\right)\sigma_{P^1\otimes P^2}^4 \underset{n\to+\infty}{\longrightarrow} 0.$$

One can prove in the same way that $\mathbb{E}\left[A_{n,2}^2\right]\underset{n\to+\infty}{\longrightarrow}0$.
Convergence of $\mathbb{E}\left[A_{n,3}^2\right]$. We easily prove that

$$\mathbb{E}\left[A_{n,3}^2\right] = \frac{1}{n^2}\sum_{1\le j<i\le n}\sum_{1\le l<k\le n}\kappa_{i,j,k,l} - \frac{1}{4n^2}\left(\mathbb{E}\left[C_{1,2}C_{2,1}\right]\right)^2,$$

where

$$\kappa_{i,j,k,l} = \mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(j)=i}\mathbb{1}_{\Pi_n(k)=l}\mathbb{1}_{\Pi_n(l)=k}\right]\times\mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{j,i}|X_j\right]\mathbb{E}\left[C_{k,l}C_{l,k}|X_l\right]\right].$$

Let us again consider $\kappa_{i,j,k,l}$ in all the cases where $1\le j<i\le n$, and $1\le l<k\le n$. If $i=k$ and $j=l$, then

$$\kappa_{i,j,k,l} = \frac{1}{n(n-1)}\mathbb{E}\left[\left(\mathbb{E}\left[C_{2,1}C_{1,2}|X_1\right]\right)^2\right].$$

If $i=k$ and $j\neq l$, or if $i\neq k$ and $j=l$, then $\kappa_{i,j,k,l}=0$.
If $i\neq k$ and $j\neq l$, then

$$\kappa_{i,j,k,l} = \frac{\left(\mathbb{E}\left[C_{1,2}C_{2,1}\right]\right)^2}{n(n-1)(n-2)(n-3)}.$$

Thus, under $(\mathcal{A}_{\varphi,Mmt})$, we finally have that

$$\mathbb{E}\left[A_{n,3}^2\right] \le \frac{1}{2n^2}\mathbb{E}\left[\left(\mathbb{E}\left[C_{1,2}C_{2,1}|X_1\right]\right)^2\right] + C\frac{n\left(\mathbb{E}\left[C_{1,2}C_{2,1}\right]\right)^2}{(n-1)(n-2)(n-3)}\underset{n\to+\infty}{\longrightarrow}0.$$

Convergence of $\mathbb{E}\left[A_{n,4}^2\right]$.

$$\mathbb{E}\left[A_{n,4}^2\right] = \frac{1}{n^2} \sum_{\substack{1\le j\neq k<i\le n \\ 1\le p\neq q<l\le n}} \Big(\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=i}\mathbb{1}_{\Pi_n(l)=p}\mathbb{1}_{\Pi_n(q)=l}\right] \times \mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{k,i}|X_j,X_k\right]\mathbb{E}\left[C_{l,p}C_{q,l}|X_p,X_q\right]\right]\Big).$$

Let us consider all the cases where $1 \le j \neq k < i \le n$, and $1 \le p \neq q < l \le n$.
If $\#\{j,k,p,q\} \ge 3$, there exists at least one element in $\{j,k,p,q\}$, j for instance (the other cases are studied in the same way), which differs from the other ones. Then,

$$\begin{aligned}
&\mathbb{E}\big[\mathbb{E}\big[C_{i,j}C_{k,i}\big|X_j,X_k\big]\mathbb{E}\left[C_{l,p}C_{q,l}|X_p,X_q\right]\big] \\
&\quad= \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{k,i}|X_j,X_k\right]\mathbb{E}\left[C_{l,p}C_{q,l}|X_p,X_q\right]|X_k,X_p,X_q\right]\right] \\
&\quad= \mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{k,i}|X_k\right]\mathbb{E}\left[C_{l,p}C_{q,l}|X_p,X_q\right]\right] \\
&\quad= \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{k,i}|X_i,X_k\right]|X_k\right]\mathbb{E}\left[C_{l,p}C_{q,l}|X_p,X_q\right]\right] \\
&\quad= \mathbb{E}\left[\mathbb{E}\left[C_{k,i}\mathbb{E}\left[C_{i,j}|X_i\right]|X_k\right]\mathbb{E}\left[C_{l,p}C_{q,l}|X_p,X_q\right]\right].
\end{aligned}$$

Since $\mathbb{E}\left[C_{i,j}|X_i\right] = 0$, this leads to

$$\mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{k,i}|X_j,X_k\right]\mathbb{E}\left[C_{l,p}C_{q,l}|X_p,X_q\right]\right] = 0. \tag{1.7.42}$$

If $j=p$, $k=q$, and $i=l$, then,

$$\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=i}\mathbb{1}_{\Pi_n(l)=p}\mathbb{1}_{\Pi_n(q)=l}\right] = \frac{1}{n(n-1)},$$

and

$$\begin{aligned}
\left|\mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{k,i}|X_j,X_k\right]\mathbb{E}\left[C_{l,p}C_{q,l}|X_p,X_q\right]\right]\right| &= \mathbb{E}\left[\left(\mathbb{E}\left[C_{i,j}C_{k,i}|X_j,X_k\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(\mathbb{E}\left[C_{3,1}C_{2,3}|X_1,X_2\right]\right)^2\right] \\
&< +\infty \quad \text{under } (\mathcal{A}_{\varphi,Mmt}).
\end{aligned}$$

If $j=p$, $k=q$, and $i\neq l$, then $\mathbb{1}_{\Pi_n(k)=i}\mathbb{1}_{\Pi_n(q)=l} = 0$, so

$$\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=i}\mathbb{1}_{\Pi_n(l)=p}\mathbb{1}_{\Pi_n(q)=l}\right] = 0.$$

If $j=q$, $k=p$, and $i=l$, then $\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(l)=p} = 0$, so

$$\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=i}\mathbb{1}_{\Pi_n(l)=p}\mathbb{1}_{\Pi_n(q)=l}\right] = 0.$$

If $j=q$, $k=p$, and $i\neq l$, then

$$\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=i}\mathbb{1}_{\Pi_n(l)=p}\mathbb{1}_{\Pi_n(q)=l}\right] = \frac{(n-4)!}{n!},$$

and

$$\begin{aligned}
\left|\mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{k,i}|X_j,X_k\right]\mathbb{E}\left[C_{l,p}C_{q,l}|X_p,X_q\right]\right]\right| &= \left|\mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{k,i}C_{l,k}C_{j,l}|X_j,X_k\right]\right]\right| \\
&\le \mathbb{E}\left[\left|C_{3,1}C_{2,3}C_{4,2}C_{1,4}\right|\right] \\
&< +\infty \quad \text{under } (\mathcal{A}_{\varphi,Mmt}).
\end{aligned}$$

By combining these results, we obtain that

$$\mathbb{E}\left[A_{n,4}^2\right] \leq C\frac{n^3}{n^2}\frac{\mathbb{E}\left[\left(\mathbb{E}\left[C_{3,1}C_{2,3}|X_1,X_2\right]\right)^2\right]}{n(n-1)} + C'\,\frac{n^4}{n^2}\frac{(n-4)!}{n!}\mathbb{E}\left[C_{3,1}C_{2,3}C_{4,2}C_{1,4}\right] \underset{n\to+\infty}{\longrightarrow} 0.$$

From (1.7.41), and the above results of convergence towards 0 for $\mathbb{E}\left[A_{n,1}^2\right]$, $\mathbb{E}\left[A_{n,2}^2\right]$, $\mathbb{E}\left[A_{n,3}^2\right]$, and $\mathbb{E}\left[A_{n,4}^2\right]$, we firstly derive that

$$\mathrm{Var}\left(\sum_{i=2}^{n}\left(\mathbb{E}\left[Y_{n,i}^2|\mathcal{F}_{n,i-1}\right]\right)\right) \underset{n\to+\infty}{\longrightarrow} 0.$$

• Second point. Notice that

$$\mathbb{E}\left[Y_{n,i}Y'_{n,i}|\mathcal{F}_{n,i-1}\right] = B_{n,1} + B_{n,2} + B_{n,3} + B_{n,4},$$

with

$$\begin{aligned}
B_{n,1} &= \frac{1}{n}\sum_{1\leq j<i\leq n}\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi'_n(i)=j}\mathbb{E}\left[C_{i,j}^2|X_j\right],\\
B_{n,2} &= \frac{1}{n}\sum_{1\leq j<i\leq n}\mathbb{1}_{\Pi_n(j)=i}\mathbb{1}_{\Pi'_n(j)=i}\mathbb{E}\left[C_{j,i}^2|X_j\right],\\
B_{n,3} &= \frac{1}{n}\sum_{1\leq j<i\leq n}\left(\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi'_n(j)=i} + \mathbb{1}_{\Pi_n(j)=i}\mathbb{1}_{\Pi'_n(i)=j}\right)\mathbb{E}\left[C_{i,j}C_{j,i}|X_j\right],
\end{aligned}$$

and

$$\begin{aligned}
B_{n,4} = \frac{1}{n}\sum_{1\leq j\neq k<i\leq n}\Big(&\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi'_n(i)=k}\mathbb{E}\left[C_{i,j}C_{i,k}|X_j,X_k\right]\\
&+\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi'_n(k)=i}\mathbb{E}\left[C_{i,j}C_{k,i}|X_j,X_k\right]\\
&+\mathbb{1}_{\Pi_n(j)=i}\mathbb{1}_{\Pi'_n(i)=k}\mathbb{E}\left[C_{j,i}C_{i,k}|X_j,X_k\right]\\
&+\mathbb{1}_{\Pi_n(j)=i}\mathbb{1}_{\Pi'_n(k)=i}\mathbb{E}\left[C_{j,i}C_{k,i}|X_j,X_k\right]\Big).
\end{aligned}$$

Thus,

$$\mathbb{E}\left[\left(\sum_{i=2}^{n}\mathbb{E}\left[Y_{n,i}Y'_{n,i}|\mathcal{F}_{n,i-1}\right]\right)^2\right] \leq 4\left(\mathbb{E}\left[B_{n,1}^2\right] + \mathbb{E}\left[B_{n,2}^2\right] + \mathbb{E}\left[B_{n,3}^2\right] + \mathbb{E}\left[B_{n,4}^2\right]\right). \quad (1.7.43)$$

Convergence of $\mathbb{E}\left[B_{n,1}^2\right]$ *and* $\mathbb{E}\left[B_{n,2}^2\right]$. It can be proved that

$$\mathbb{E}\left[B_{n,1}^2\right] \leq \frac{1}{n^3}\sum_{1\leq j<i\leq n}\sum_{1\leq l<k\leq n}\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=l}\right]\times\mathbb{E}\left[\mathbb{E}\left[C_{i,j}^2|X_j\right]\mathbb{E}\left[C_{k,l}^2|X_l\right]\right].$$

Then, with the same computations as for the convergence of $\mathbb{E}\left[A_{n,1}^2\right]$ above, we prove that

$$\mathbb{E}\left[B_{n,1}^2\right] \leq \frac{n-1}{2n^3}\mathbb{E}\left[\left(\mathbb{E}\left[C_{1,2}^2|X_2\right]\right)^2\right] + C\frac{\sigma_{P^1\otimes P^2}^4}{n-1} \underset{n\to+\infty}{\longrightarrow} 0.$$

In the same way, we also prove that $\mathbb{E}\left[B_{n,2}^2\right] \underset{n\to+\infty}{\longrightarrow} 0$.

Convergence of $\mathbb{E}\left[B_{n,3}^2\right]$. We also have that

$$\mathbb{E}\left[B_{n,3}^2\right] \leq \frac{4}{n^2} \sum_{1\leq j<i\leq n} \sum_{1\leq l<k\leq n} \mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=l}\right] \times \\ \mathbb{E}\left[\mathbb{1}_{\Pi'_n(j)=i}\mathbb{1}_{\Pi'_n(l)=k}\right] \mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{j,i}|X_j\right]\mathbb{E}\left[C_{k,l}C_{l,k}|X_l\right]\right].$$

Now, with similar computations as for the convergence of $\mathbb{E}\left[A_{n,1}^2\right]$ above again, we prove that

$$\mathbb{E}\left[B_{n,3}^2\right] \quad \leq \quad 2\frac{n-1}{n^3}\mathbb{E}\left[\left(\mathbb{E}\left[C_{1,2}C_{2,1}|X_2\right]\right)^2\right] + C\frac{\left(\mathbb{E}\left[C_{1,2}C_{2,1}\right]\right)^2}{n-1} \underset{n\to+\infty}{\longrightarrow} 0.$$

Convergence of $\mathbb{E}\left[B_{n,4}^2\right]$. Setting

$$B_{n,4,1} = \frac{1}{n}\sum_{1\leq j\neq k<i\leq n} \mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi'_n(i)=k}\mathbb{E}\left[C_{i,j}C_{i,k}|X_j,X_k\right],$$

$$B_{n,4,2} = \frac{1}{n}\sum_{1\leq j\neq k<i\leq n} \mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi'_n(k)=i}\mathbb{E}\left[C_{i,j}C_{k,i}|X_j,X_k\right],$$

$$B_{n,4,3} = \frac{1}{n}\sum_{1\leq j\neq k<i\leq n} \mathbb{1}_{\Pi_n(j)=i}\mathbb{1}_{\Pi'_n(i)=k}\mathbb{E}\left[C_{j,i}C_{i,k}|X_j,X_k\right],$$

$$B_{n,4,4} = \frac{1}{n}\sum_{1\leq j\neq k<i\leq n} \mathbb{1}_{\Pi_n(j)=i}\mathbb{1}_{\Pi'_n(k)=i}\mathbb{E}\left[C_{j,i}C_{k,i}|X_j,X_k\right],$$

then $B_{n,4} = B_{n,4,1} + B_{n,4,2} + B_{n,4,3} + B_{n,4,4}$ and in particular,

$$\mathbb{E}\left[{B_{n,4}}^2\right] \leq 4\left(\mathbb{E}\left[{B_{n,4}}^2\right] + \mathbb{E}\left[{B_{n,4}}^2\right] + \mathbb{E}\left[{B_{n,4}}^2\right] + \mathbb{E}\left[{B_{n,4}}^2\right]\right).$$

Yet,

$$\mathbb{E}\left[B_{n,4,1}^2\right] = \frac{1}{n^2} \sum_{1\leq j\neq k<i\leq n} \sum_{1\leq p\neq q<l\leq n} \mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(l)=p}\right] \times \\ \mathbb{E}\left[\mathbb{1}_{\Pi'_n(i)=k}\mathbb{1}_{\Pi'_n(l)=q}\right] \mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{i,k}|X_j,X_k\right]\mathbb{E}\left[C_{l,p}C_{l,q}|X_p,X_q\right]\right].$$

Now, consider all the cases where $1 \leq j \neq k < i \leq n$, $1 \leq p \neq q < l \leq n$.
If $\#\{j,k,p,q\} \geq 3$, using a similar argument as in (1.7.42), we obtain that

$$\mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{i,k}|X_j,X_k\right]\mathbb{E}\left[C_{l,p}C_{l,q}|X_p,X_q\right]\right] = 0.$$

If $j=p$, $k=q$, and $i=l$, then,

$$\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(l)=p}\right]\mathbb{E}\left[\mathbb{1}_{\Pi'_n(i)=k}\mathbb{1}_{\Pi'_n(l)=q}\right] = \frac{1}{n^2},$$

and

$$\begin{aligned}\left|\mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{i,k}|X_j,X_k\right]\mathbb{E}\left[C_{l,p}C_{l,q}|X_p,X_q\right]\right]\right| &= \mathbb{E}\left[\left(\mathbb{E}\left[C_{i,j}C_{i,k}|X_j,X_k\right]\right)^2\right] \\ &= \mathbb{E}\left[\left(\mathbb{E}\left[C_{3,1}C_{3,2}|X_1,X_2\right]\right)^2\right] \\ &< +\infty \quad \text{under } (\mathcal{A}_{\varphi,Mmt}).\end{aligned}$$

If $j=p$, $k=q$, $i\neq l$, or if $j=q$, $k=p$, $i=l$, then $\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(l)=p}$ is equal to 0, so

$$\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(l)=p}\right]\mathbb{E}\left[\mathbb{1}_{\Pi'_n(i)=k}\mathbb{1}_{\Pi'_n(l)=q}\right]=0.$$

If $j=q$, $k=p$, and $i\neq l$, then

$$\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(l)=p}\right]\mathbb{E}\left[\mathbb{1}_{\Pi'_n(i)=k}\mathbb{1}_{\Pi'_n(l)=q}\right]=\frac{1}{n^2(n-1)^2},$$

and

$$\begin{aligned}\left|\mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{i,k}|X_j,X_k\right]\mathbb{E}\left[C_{l,p}C_{l,q}|X_p,X_q\right]\right]\right| &= \left|\mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{i,k}C_{l,k}C_{l,j}|X_j,X_k\right]\right]\right|\\ &= \mathbb{E}\left[|C_{3,1}C_{2,3}C_{4,2}C_{1,4}|\right]\\ &< +\infty \quad \text{under } (\mathcal{A}_{\varphi,Mmt}).\end{aligned}$$

By combining these results, we obtain that

$$\mathbb{E}\left[B_{n,4,1}^2\right] \leq C\frac{\mathbb{E}\left[\left(\mathbb{E}\left[C_{3,1}C_{3,2}|X_1,X_2\right]\right)^2\right]}{n}+C'\frac{\mathbb{E}\left[C_{3,1}C_{2,3}C_{4,2}C_{1,4}\right]}{(n-1)^2}\underset{n\to+\infty}{\longrightarrow}0.$$

Following the same lines of proof, we furthermore obtain that $\mathbb{E}\left[B_{n,4,2}^2\right]$, $\mathbb{E}\left[B_{n,4,3}^2\right]$, and $\mathbb{E}\left[B_{n,4,4}^2\right]$ also converge towards 0. Hence, $\mathbb{E}\left[B_{n,4}^2\right]\underset{n\to+\infty}{\longrightarrow}0$. From (1.7.43), and the convergence towards 0 of $\mathbb{E}\left[B_{n,1}^2\right]$, $\mathbb{E}\left[B_{n,2}^2\right]$, $\mathbb{E}\left[B_{n,3}^2\right]$, and $\mathbb{E}\left[B_{n,4}^2\right]$, we derive that

$$\mathbb{E}\left[\left(\sum_{i=2}^{n}\mathbb{E}\left[Y_{n,i}Y'_{n,i}|\mathcal{F}_{n,i-1}\right]\right)^2\right]\underset{n\to+\infty}{\longrightarrow}0,$$

which finally allows to conclude that assumption (i) is satisfied.

Assumption (ii)**.** Given $\varepsilon>0$, let us prove that

$$\sum_{i=2}^{n}\mathbb{E}\left[\left(aY_{n,i}+bY'_{n,i}\right)^2\mathbb{1}_{|aY_{n,i}+bY'_{n,i}|>\varepsilon}\right]\underset{n\to+\infty}{\longrightarrow}0.$$

$$\begin{aligned}\sum_{i=2}^{n}\mathbb{E}\left[\left(aY_{n,i}+bY'_{n,i}\right)^2\mathbb{1}_{|aY_{n,i}+bY'_{n,i}|>\varepsilon}\right] &\leq \frac{1}{\varepsilon^2}\sum_{i=2}^{n}\mathbb{E}\left[\left(aY_{n,i}+bY'_{n,i}\right)^4\right]\\ &\leq \frac{2^3}{\varepsilon^2}\sum_{i=2}^{n}\left(a^4\mathbb{E}\left[Y_{n,i}^4\right]+b^4\mathbb{E}\left[{Y'_{n,i}}^4\right]\right)\\ &\leq \frac{2^3(a^4+b^4)}{\varepsilon^2}\sum_{i=2}^{n}\mathbb{E}\left[Y_{n,i}^4\right].\end{aligned}$$

Since $Y_{n,i}=n^{-1/2}\left(\mathbb{1}_{\Pi_n(i)<i}C_{i,\Pi_n(i)}+\mathbb{1}_{\Pi_n^{-1}(i)<i}C_{\Pi_n^{-1}(i),i}\right)$,

$$\begin{aligned}\mathbb{E}\left[Y_{n,i}^4\right] &\leq \frac{2^3}{n^2}\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)<i}C_{i,\Pi_n(i)}^4+\mathbb{1}_{\Pi_n^{-1}(i)<i}C_{\Pi_n^{-1}(i),i}^4\right]\\ &\leq \frac{2^3}{n^2}\sum_{j=1}^{i-1}\left(\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}C_{i,j}^4\right]+\mathbb{E}\left[\mathbb{1}_{\Pi_n^{-1}(i)=j}C_{j,i}^4\right]\right)\\ &\leq \frac{2^4}{n^2}\mathbb{E}\left[C_{1,2}^4\right].\end{aligned}$$

We thus obtain that

$$\sum_{i=2}^{n} \mathbb{E}\left[\left(aY_{n,i}+bY'_{n,i}\right)^2 \mathbb{1}_{\left|aY_{n,i}+bY'_{n,i}\right|>\varepsilon}\right] \leq \frac{2^7(a^4+b^4)}{\varepsilon^2 n}\mathbb{E}\left[C_{1,2}^4\right],$$

where the right-hand side tends to 0 as soon as $\mathbb{E}\left[C_{1,2}^4\right] < +\infty$.
This last condition is ensured by $(\mathcal{A}_{\varphi,Mmt})$, which allows to confirm that assumption (ii) is also checked, and that

$$\mathcal{L}\left(aM_n^{\Pi_n}(\mathbb{X}_n)+bM_n^{\Pi'_n}(\mathbb{X}_n)\right) \underset{n\to+\infty}{\Longrightarrow} \mathcal{N}\left(0,\left(a^2+b^2\right)\sigma^2_{P^1\otimes P^2}\right).$$

This ends the proof of Lemma 1.7.1.

Recall that we aim at proving that

$$d_{BL}\left(\mathcal{L}\left(M_n^{\Pi_n}(\mathbb{X}_n)|\mathbb{X}_n\right),\mathcal{N}\left(0,\sigma^2_{P^1\otimes P^2}\right)\right) \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

From Lemma 1.7.1, we deduce that for every t in $\mathbb{R}$,

$$\begin{cases} \mathbb{P}\left(M_n^{\Pi_n}(\mathbb{X}_n)\leq t\right) \underset{n\to+\infty}{\longrightarrow} \Phi_{0,\sigma^2_{P^1\otimes P^2}}(t), \\ \mathbb{P}\left(M_n^{\Pi_n}(\mathbb{X}_n)\leq t, M_n^{\Pi'_n}(\mathbb{X}_n)\leq t\right) \underset{n\to+\infty}{\longrightarrow} \Phi^2_{0,\sigma^2_{P^1\otimes P^2}}(t). \end{cases}$$

Setting $M_n = M_n^{\Pi_n}(\mathbb{X}_n)$ for the sake of simplicity, this leads to

$$\begin{cases} \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{M_n\leq t}|\mathbb{X}_n\right]\right] \underset{n\to+\infty}{\longrightarrow} \Phi_{0,\sigma^2_{P^1\otimes P^2}}(t), \\ \mathbb{E}\left[\left(\mathbb{E}\left[\mathbb{1}_{M_n\leq t}|\mathbb{X}_n\right]\right)^2\right] \underset{n\to+\infty}{\longrightarrow} \Phi^2_{0,\sigma^2_{P^1\otimes P^2}}(t). \end{cases} \tag{1.7.44}$$

In a separable metric space, convergence in probability is metrizable, therefore it is equivalent to almost sure convergence of a subsequence of any initial subsequence (see, e.g., [46, Th. 9.2.1] recalled in Appendix A.1, Theorem A.1.1). So, let us fix an initial extraction $\phi_0 : \mathbb{N}\to\mathbb{N}$, which defines a subsequence $\left(M_{\phi_0(n)}\right)_{n\in\mathbb{N}}$ of $(M_n)_{n\in\mathbb{N}}$. Let us denote by $(q_m)_{m\in\mathbb{N}}$ a sequence such that $\{q_m, m\in\mathbb{N}\}=\mathbb{Q}$. For any m in $\mathbb{N}$, from (1.7.44), we derive that

$$\begin{cases} \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{M_{\phi_0(n)}\leq q_m}\middle|\mathbb{X}_{\phi_0(n)}\right]\right] \underset{n\to+\infty}{\longrightarrow} \Phi_{0,\sigma^2_{P^1\otimes P^2}}(q_m), \\ \mathbb{E}\left[\left(\mathbb{E}\left[\mathbb{1}_{M_{\phi_0(n)}\leq q_m}\middle|\mathbb{X}_{\phi_0(n)}\right]^2\right)\right] \underset{n\to+\infty}{\longrightarrow} \Phi^2_{0,\sigma^2_{P^1\otimes P^2}}(q_m), \end{cases}$$

which leads (by Chebychev's inequality) to

$$\mathbb{E}\left[\mathbb{1}_{M_{\phi_0(n)}\leq q_m}\middle|\mathbb{X}_{\phi_0(n)}\right] \xrightarrow[n\to+\infty]{\mathbb{P}} \Phi_{0,\sigma^2_{P^1\otimes P^2}}(q_m). \tag{1.7.45}$$

Therefore, there exist an extraction ϕ_1 and a subset Ω_1 of Ω such that $\mathbb{P}(\Omega_1)=1$, and for every ω in Ω_1,

$$\mathbb{E}\left[\mathbb{1}_{M_{\phi_1\circ\phi(n)}\leq q_1}\middle|\mathbb{X}_{\phi_1\circ\phi(n)}\right](\omega) \underset{n\to+\infty}{\longrightarrow} \Phi_{0,\sigma^2_{P^1\otimes P^2}}(q_1).$$

Now, let $m\geq 1$ for which there exist an extraction ϕ_m and a subset Ω_m of Ω such that $\mathbb{P}(\Omega_m)=1$, and for every $\omega\in\Omega_m$,

$$\mathbb{E}\left[\mathbb{1}_{M_{\phi_m\circ\phi_{m-1}\circ\ldots\circ\phi_0(n)}\leq q_m}\middle|\mathbb{X}_{\phi_m\circ\phi_{m-1}\circ\ldots\circ\phi_0(n)}\right](\omega) \underset{n\to+\infty}{\longrightarrow} \Phi_{0,\sigma^2_{P^1\otimes P^2}}(q_m).$$

Then, from (1.7.45), there also exist an extraction ϕ_{m+1} and a subset Ω_{m+1} of Ω such that $\mathbb{P}(\Omega_{m+1}) = 1$, and for every ω in Ω_{m+1},

$$\mathbb{E}\left[\mathbb{1}_{M_{\phi_{m+1}\circ\phi_m\circ\phi_{m-1}\circ\ldots\circ\phi_0(n)}\leq q_{m+1}}\middle|\mathbb{X}_{\phi_{m+1}\circ\phi_m\circ\ldots\circ\phi_0(n)}\right](\omega) \underset{n\to+\infty}{\longrightarrow} \Phi_{0,\sigma^2_{P^1\otimes P^2}}(q_{m+1}).$$

Setting $\tilde{\Omega} = \bigcap_{m\in\mathbb{N}}\Omega_m$ and for every n in $\mathbb{N}$, $\tilde{\phi}(n) = \phi_n\circ\ldots\circ\phi_2\circ\phi_1(n)$, then $\mathbb{P}\left(\tilde{\Omega}\right) = 1$. Moreover, for every ω in $\tilde{\Omega}$, m in $\mathbb{N}$,

$$\mathbb{E}\left[\mathbb{1}_{M_{\tilde{\phi}\circ\phi_0(n)}\leq q_m}\middle|\mathbb{X}_{\tilde{\phi}\circ\phi_0(n)}\right](\omega) \underset{n\to+\infty}{\longrightarrow} \Phi_{0,\sigma^2_{P^1\otimes P^2}}(q_m).$$

Since $\Phi_{0,\sigma^2_{P^1\otimes P^2}}$ is a continuous distribution function, it can be proved that this follows

$$d_{BL}\left(\mathcal{L}\left(M_{\tilde{\phi}\circ\phi_0(n)}\middle|\mathbb{X}_{\tilde{\phi}\circ\phi_0(n)}\right),\mathcal{N}\left(0,\sigma^2_{P^1\otimes P^2}\right)\right) \underset{n\to+\infty}{\overset{a.s.}{\longrightarrow}} 0.$$

To conclude, we actually proved that

$$d_{BL}\left(\mathcal{L}\left(M_n^{\Pi_n}(\mathbb{X}_n)\middle|\mathbb{X}_n\right),\mathcal{N}\left(0,\sigma^2_{P^1\otimes P^2}\right)\right) \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0,$$

which, combined with (1.7.36), leads to

$$d_{BL}\left(\mathcal{L}\left(\sqrt{n}U_n\left(\mathbb{X}_n^{\Pi_n}\right)\middle|\mathbb{X}_n\right),\mathcal{N}\left(0,\sigma^2_{P^1\otimes P^2}\right)\right) \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0.$$

Third step: convergence of conditional second-order moments. Recall that from (1.7.34), $U_n\left(\mathbb{X}_n^{\Pi_n}\right) = \dfrac{1}{n-1}U_n^{\Pi_n}$, where

$$\begin{aligned} U_n^{\Pi_n} &= \sum_{i=1}^n \varphi\left(X_i^1, X^2_{\Pi_n(i)}\right) - \frac{1}{n}\sum_{i,j=1}^n \varphi\left(X_i^1, X_j^2\right) \\ &= \sum_{i,j=1}^n \left(\mathbb{1}_{\Pi_n(i)=j} - \frac{1}{n}\right)\varphi\left(X_i^1, X_j^2\right). \end{aligned}$$

Therefore,

$$\mathbb{E}\left[\left(\sqrt{n}U_n\left(\mathbb{X}_n^{\Pi_n}\right)\right)^2\middle|\mathbb{X}_n\right] = \frac{n^2}{(n-1)^2}\left(\frac{1}{n}\mathbb{E}\left[\left(U_n^{\Pi_n}\right)^2\middle|\mathbb{X}_n\right]\right), \tag{1.7.46}$$

and if $C_{i,j,k,l} = \left(\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=l}\right] - \frac{1}{n^2}\right)\varphi\left(X_i^1, X_j^2\right)\varphi\left(X_k^1, X_l^2\right)$, then

$$\frac{1}{n}\mathbb{E}\left[\left(U_n^{\Pi_n}\right)^2\middle|\mathbb{X}_n\right] = \frac{1}{n}\sum_{i,j=1}^n\sum_{k,l=1}^n C_{i,j,k,l}.$$

Firstly,

$$\frac{1}{n}\sum_{\substack{i,j,k,l\in\{1,\ldots,n\}\\ \#\{i,j,k,l\}=4}} C_{i,j,k,l} = \frac{(n-2)(n-3)}{n^2}U_{n,1},$$

where

$$U_{n,1} = \frac{(n-4)!}{n!}\sum_{\substack{i,j,k,l\in\{1,\ldots,n\}\\ \#\{i,j,k,l\}=4}} \varphi\left(X_i^1, X_j^2\right)\varphi\left(X_k^1, X_l^2\right)$$

is clearly a U-statistic of order 4. From the strong law of large numbers of Hœffding [85] (see Appendix A.3.2, Theorem A.3.1), we thus have that

$$\frac{(n-2)(n-3)}{n^2}U_{n,1} \xrightarrow[n\to+\infty]{a.s.} \left(\mathbb{E}\left[\varphi\left(X_1^1, X_2^2\right)\right]\right)^2.$$

Secondly,

$$\frac{1}{n}\sum_{\substack{i,j,k,l\in\{1,\ldots,n\}\\ \#\{i,j,k,l\}=3\\ i=j, i=l, j=k, \text{ or } k=l}} C_{i,j,k,l} = \frac{2(n-2)}{n^2}U_{n,2},$$

where

$$U_{n,2} = \frac{(n-3)!}{n!}\sum_{\substack{i,k,l\in\{1,\ldots,n\}\\ \#\{i,k,l\}=3}} \left(\varphi\left(X_i^1, X_i^2\right)\varphi\left(X_k^1, X_l^2\right) + \varphi\left(X_i^1, X_l^2\right)\varphi\left(X_k^1, X_i^2\right)\right)$$

is a U-statistic of order 3 which converges almost surely, so

$$\frac{2(n-2)}{n^2}U_{n,2} \xrightarrow[n\to+\infty]{a.s.} 0.$$

Thirdly,

$$\frac{1}{n}\sum_{\substack{i,j,k,l\in\{1,\ldots,n\}\\ \#\{i,j,k,l\}=3\\ i=k, \text{ or } j=l}} C_{i,j,k,l} = -\frac{n(n-1)(n-2)}{n^3}U_{n,3},$$

where

$$U_{n,3} = \frac{(n-3)!}{n!}\sum_{\substack{i,k,l\in\{1,\ldots,n\}\\ \#\{i,k,l\}=3}} \left(\varphi\left(X_i^1, X_k^2\right)\varphi\left(X_i^1, X_l^2\right) + \varphi\left(X_i^1, X_l^2\right)\varphi\left(X_k^1, X_l^2\right)\right)$$

is a U-statistic of order 3. So,

$$-\frac{n(n-1)(n-2)}{n^3}U_{n,3} \xrightarrow[n\to+\infty]{a.s.} -\mathbb{E}\left[\left(\mathbb{E}\left[\varphi(X_1^1, X_2^2)\middle|X_1\right]\right)^2\right] - \mathbb{E}\left[\left(\mathbb{E}\left[\varphi(X_1^1, X_2^2)\middle|X_2\right]\right)^2\right].$$

Fourthly,

$$\frac{1}{n}\sum_{\substack{i,j,k,l\in\{1,\ldots,n\}\\ \#\{i,j,k,l\}=2\\ i=j=k, i=j=l,\\ i=k=l, \text{ or } j=k=l}} C_{i,j,k,l} = -\frac{2(n-1)}{n^2}U_{n,4},$$

where

$$U_{n,4} = \frac{1}{n(n-1)}\sum_{1\le i\neq j\le n} \left(\varphi\left(X_i^1, X_i^2\right)\varphi\left(X_i^1, X_j^2\right) + \varphi\left(X_i^1, X_i^2\right)\varphi\left(X_j^1, X_i^2\right)\right)$$

is a U-statistic of order 2, so

$$-\frac{2(n-1)}{n^2}U_{n,4} \xrightarrow[n\to+\infty]{a.s.} 0.$$

Fifthly,

$$\frac{1}{n} \sum_{\substack{i,j,k,l \in \{1,\dots,n\} \\ \#\{i,j,k,l\}=2 \\ i=j\neq k=l, \text{ or } i=l\neq j=k}} C_{i,j,k,l} = \frac{1}{n^2} U_{n,5},$$

where

$$U_{n,5} = \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} \left(\varphi\left(X_i^1, X_i^2\right) \varphi\left(X_j^1, X_j^2\right) + \varphi\left(X_i^1, X_j^2\right) \varphi\left(X_j^1, X_i^2\right)\right)$$

is a U-statistic of order 2, so

$$\frac{1}{n^2} U_{n,5} \xrightarrow[n\to+\infty]{a.s.} 0.$$

Sixthly,

$$\frac{1}{n} \sum_{\substack{i,j,k,l \in \{1,\dots,n\} \\ \#\{i,j,k,l\}=2 \\ i=k\neq j=l}} C_{i,j,k,l} = \frac{(n-1)^2}{n^2} U_{n,6},$$

where

$$U_{n,6} = \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} \varphi^2\left(X_i^1, X_j^2\right)$$

is a U-statistic of order 2, so

$$\frac{(n-1)^2}{n^2} U_{n,6} \xrightarrow[n\to+\infty]{a.s.} \mathbb{E}\left[\varphi^2\left(X_1^1, X_2^2\right)\right].$$

Seventhly,

$$\frac{1}{n} \sum_{\substack{i,j,k,l \in \{1,\dots,n\} \\ \#\{i,j,k,l\}=1}} C_{i,j,k,l} = \frac{n-1}{n^3} \sum_{i=1}^{n} \varphi\left(X_i^1, X_i^2\right),$$

which almost surely tends to 0 thanks to the strong law of large numbers.
By combining all these results, and the fact that

$$\sigma^2_{P^1\otimes P^2} = \mathbb{E}\left[\varphi^2\left(X_1^1, X_2^2\right)\right] + \left(\mathbb{E}\left[\varphi\left(X_1^1, X_2^2\right)\right]\right)^2 - \mathbb{E}\left[\left(\mathbb{E}\left[\varphi(X_1^1, X_2^2)|X_1\right]\right)^2\right] - \mathbb{E}\left[\left(\mathbb{E}\left[\varphi(X_1^1, X_2^2)|X_2\right]\right)^2\right],$$

we finally obtain that

$$\frac{1}{n}\mathbb{E}\left[\left(U_n^{\Pi_n}\right)^2 \middle| \mathbb{X}_n\right] \xrightarrow[n\to+\infty]{a.s.} \sigma^2_{P^1\otimes P^2},$$

and from (1.7.46), we deduce that (1.7.33) is satisfied, that is

$$\mathbb{E}\left[\left(\sqrt{n} U_n\left(\mathbb{X}_n^{\Pi_n}\right)\right)^2 \middle| \mathbb{X}_n\right] \xrightarrow[n\to+\infty]{a.s.} \sigma^2_{P^1\otimes P^2}.$$

Since $d_{BL}\left(\mathcal{L}\left(\sqrt{n} U_n\left(\mathbb{X}_n^{\Pi_n}\right)|\mathbb{X}_n\right), \mathcal{N}\left(0, \sigma^2_{P^1\otimes P^2}\right)\right) \xrightarrow[n\to+\infty]{\mathbb{P}} 0$, this allows to conclude that

$$d_2\left(\mathcal{L}\left(\sqrt{n} U_n\left(\mathbb{X}_n^{\Pi_n}\right)|\mathbb{X}_n\right), \mathcal{N}\left(0, \sigma^2_{P^1\otimes P^2}\right)\right) \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

1.7.12 Proof of Corollary 1.4.1

Here, unlike the bootstrap approach, we only have in Theorem 1.4.1 a consistency result in probability. Thus, as for Proposition 1.3.6, we use an argument of subsequences. So let $\phi_0 : \mathbb{N} \to \mathbb{N}$ be an extraction defining a subsequence. By Theorem 1.4.1, there exists an extraction ϕ_1 such that P-a.s. in $(X_i)_i$,

$$\mathcal{L}\left(\sqrt{\phi_1\circ\phi_0(n)}U_{\phi_1\circ\phi_0(n)}\left(\mathbb{X}^{\Pi_{\phi_1\circ\phi_0(n)}}_{\phi_1\circ\phi_0(n)}\right)\middle|\mathbb{X}_{\phi_1\circ\phi_0(n)}\right) \underset{n\to+\infty}{\Longrightarrow} \mathcal{N}\left(0,\sigma^2_{P^1\otimes P^2}\right). \tag{1.7.47}$$

In particular, applying [172, Lemma 21.2] (see Appendix A.1, Lemma A.1.2) on the event where the convergence is true, we obtain that for η in $(0,1)$,

$$q^\star_{\eta,\phi_1\circ\phi_0(n)}\left(\mathbb{X}_{\phi_1\circ\phi_0(n)}\right) \underset{n\to+\infty}{\overset{a.s.}{\longrightarrow}} \Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(\eta),$$

which ends the proof by [46, Th. 9.2.1] recalled in Appendix A.1, Theorem A.1.1.

1.7.13 Proof of Theorem 1.4.2

The proof of Theorem 1.4.2 for the tests of $\Gamma(q^\star)$ is very similar to the one of Theorem 1.3.2, just replacing the argument of (1.7.27) by

$$q^\star_{1-\alpha,n}\left(\mathbb{X}_n\right) \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} \Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(1-\alpha),$$

which is derived from Corollary 1.4.1.
Now for the tests with a Monte Carlo approximation of the quantiles, we use arguments similar to those of Proposition 1.3.6, still focusing on the upper-tailed tests of $\Gamma(q^\star_{MC})$. We therefore aim here at proving that

$$\sqrt{n}U^{\star(\lceil(1-\alpha)(B_n+1)\rceil)} \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} \Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(1-\alpha). \tag{1.7.48}$$

Then, one can conclude as in the proof of Theorem 1.3.2.
Let $F^\star_{n,\mathbb{X}_n}$ be the c.d.f. of $\mathcal{L}\left(\sqrt{n}U_n, P^\star_n|\mathbb{X}_n\right)$, and let us first prove that

$$\sup_{z\in\mathbb{R}}\left|F^\star_{n,\mathbb{X}_n}(z) - \Phi_{0,\sigma^2_{P^1\otimes P^2}}(z)\right| \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0. \tag{1.7.49}$$

As Theorem 1.4.1 provides only a convergence in probability, similar arguments of subsequences as in the proof of Corollary 1.4.1, have to be used. So, let ϕ_0 be an initial extraction and ϕ_1 be the extraction such that (1.7.47) is satisfied. As convergence in the d_{BL} metric is equivalent to a weak convergence (see [46, Theorem 11.3.3] recalled in Appendix A.1.2), and as the limit is continuous, by [172, Lemma 2.11] (see Appendix A.1, Lemma A.1.1) we obtain that

$$\sup_{z\in\mathbb{R}}\left|F^\star_{\phi_1\circ\phi_0(n),\mathbb{X}_{\phi_1\circ\phi_0(n)}}(z) - \Phi_{0,\sigma^2_{P^1\otimes P^2}}(z)\right| \underset{n\to+\infty}{\overset{a.s.}{\longrightarrow}} 0.$$

This being true for any initial extraction ϕ_0, we obtain (1.7.49).
Let $F^{\star B_n}_{n,\mathbb{X}_n}$ denote the empirical c.d.f. of $\mathcal{L}\left(\sqrt{n}U_n, P^\star_n|\mathbb{X}_n\right)$ associated with the sample $\left(\Pi^1_n,\ldots,\Pi^{B_n}_n\right)$, that is

$$\forall z\in\mathbb{R},\quad F^{\star B_n}_{n,\mathbb{X}_n}(z) = \frac{1}{B_n}\sum_{b=1}^{B_n}\mathbb{1}_{\sqrt{n}U_n\left(\mathbb{X}^{\Pi^b_n}_n\right)\le z}.$$

Then, by the Dvoretzky-Kiefer-Wolfowitz inequality (see [122, Corollary 1] or Appendix A.1.2, Theorem A.1.5), we obtain as in the proof of Proposition 1.3.6,

$$\sup_{z\in\mathbb{R}}\left|F^{\star B_n}_{n,\mathbb{X}_n}(z)-F^{\star}_{n,\mathbb{X}_n}(z)\right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0. \tag{1.7.50}$$

Finally, let

$$G^{\star B_n}_{n,\mathbb{X}_n}(z)=\frac{1}{B_n+1}\sum_{b=1}^{B_n+1}\mathbb{1}_{\sqrt{n}U^{\star b}\le z}.$$

Since $G^{\star B_n}_{n,\mathbb{X}_n}(z)=\frac{1}{B_n+1}\left(\mathbb{1}_{\sqrt{n}U_n(\mathbb{X}_n)\le z}+B_nF^{\star B_n}_{n,\mathbb{X}_n}(z)\right)$,

$$\sup_{z\in\mathbb{R}}\left|G^{\star B_n}_{n,\mathbb{X}_n}(z)-F^{\star B_n}_{n,\mathbb{X}_n}(z)\right|\le\frac{2}{B_n+1}\underset{n\to+\infty}{\longrightarrow}0. \tag{1.7.51}$$

Combining (1.7.49), (1.7.50) and (1.7.51) leads to:

$$\sup_{z\in\mathbb{R}}\left|G^{\star B_n}_{n,\mathbb{X}_n}(z)-\Phi_{0,\sigma^2_{P^1\otimes P^2}}(z)\right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Since

$$\sqrt{n}U^{\star(\lceil(1-\alpha)(B_n+1)\rceil)}=\left(G^{\star B_n}_{n,\mathbb{X}_n}\right)^{-1}(1-\alpha),$$

we obtain (1.7.48), and as explained above, we conclude as in the proof of Theorem 1.3.2.

1.8 Additional Results

1.8.1 About the non-degeneracy of the U-statistic

We focus on the *Linear case* with $\varphi=\varphi^w$ given by (1.2.2). Define

$$\begin{aligned}Z(x)=&\int_{[0,1]^2}w(u,v)dN_{x^1}(u)dN_{x^2}(v)+\mathbb{E}\left[\int_{[0,1]^2}w(u,v)dN_{X^1}(u)dN_{X^2}(v)\right]\\&-\mathbb{E}\left[\int_{[0,1]^2}w(u,v)dN_{x^1}(u)dN_{X^2}(v)\right]-\mathbb{E}\left[\int_{[0,1]^2}w(u,v)dN_{X^1}(u)dN_{x^2}(v)\right].\end{aligned}$$

Recall that in this case, degeneracy is equivalent to stating that for $X=(X^1,X^2)$ with distribution $P^1\otimes P^2$, $Z(X)$ is a random variable which is almost surely null under $(\mathcal{H}_0)$. Since $\mathbb{E}[Z(X)]=0$, $Z(X)=0$ a.s. is equivalent to $\mathrm{Var}(Z(X))=0$. Here we provide a computation of $\mathrm{Var}(Z(X))$.
Let us introduce $dM_1^{[1]}(u)$ and $dM_2^{[1]}(v)$ the mean measures of respectively X^1 with distribution P^1 and X^2 of distribution P^2 [38, Chapter 5], then one can rewrite

$$\begin{aligned}Z(X)=&\int_{[0,1]^2}w(u,v)dN_{X^1}(u)dN_{X^2}(v)+\int_{[0,1]^2}w(u,v)dM_1^{[1]}(u)dM_2^{[1]}(v)\\&-\int_{[0,1]^2}w(u,v)dN_{X^1}(u)dM_2^{[1]}(v)-\int_{[0,1]^2}w(u,v)dM_1^{[1]}(u)dN_{X^2}(v).\end{aligned}$$

Therefore, $\mathbb{E}\left[Z(X)\right] = 0$, and

$$\begin{aligned}
\mathrm{Var}(Z(X)) &= \mathbb{E}\left[Z(X)^2\right] \\
&= \int_{[0,1]^4} w(u,v)w(s,t)\mathbb{E}\left[dN_{X^1}(u)dN_{X^1}(s)\right]\mathbb{E}\left[dN_{X^2}(v)dN_{X^2}(t)\right] \\
&\quad - \int_{[0,1]^4} w(u,v)w(s,t)\mathbb{E}\left[dN_{X^1}(u)dN_{X^1}(s)\right] dM_2^{[1]}(v)dM_2^{[1]}(t) \\
&\quad - \int_{[0,1]^4} w(u,v)w(s,t)dM_1^{[1]}(u)dM_1^{[1]}(s)\mathbb{E}\left[dN_{X^2}(v)dN_{X^2}(t)\right] \\
&\quad + \int_{[0,1]^4} w(u,v)w(s,t)dM_1^{[1]}(u)dM_1^{[1]}(s)dM_2^{[1]}(v)dM_2^{[1]}(t).
\end{aligned}$$

By assuming that $\#X^1$ (resp. $\#X^2$) has second-order moment, (see also Section 1.3.2 for comment on this assumption), one can introduce the second factorial moment measure associated with X^1 (resp. X^2), and denoted by $dM_1^{[2]}(u,s)$ (resp. $dM_2^{[2]}(v,t)$). Then straightforward computations show that

$$\begin{aligned}
\mathrm{Var}(Z(X)) = &\int_{[0,1]^2} w(u,v)^2 dM_1^{[1]}(u)dM_2^{[1]}(v) \\
&+ \int_{[0,1]^3} w(u,v)w(u,t)dM_1^{[1]}(u)\left(dM_2^{[2]}(v,t) - dM_2^{[1]}(v)dM_2^{[1]}(t)\right) \\
&+ \int_{[0,1]^3} w(u,v)w(s,v)\left(dM_1^{[2]}(u,s) - dM_1^{[1]}(u)dM_1^{[1]}(s)\right) dM_2^{[1]}(v) \\
&+ \int_{[0,1]^4} w(u,v)w(s,t)\left(dM_1^{[2]}(u,s) - dM_1^{[1]}(u)dM_1^{[1]}(s)\right) \times \\
&\qquad\qquad \left(dM_2^{[2]}(v,t) - dM_2^{[1]}(v)dM_2^{[1]}(t)\right).
\end{aligned}$$

In particular, for Poisson processes, $dM^{[2]}(u,s) = dM^{[1]}(u)dM^{[1]}(s)$ and

$$\mathrm{Var}(Z(X)) = \int_{[0,1]^2} w(u,v)^2 dM_1^{[1]}(u)dM_2^{[1]}(v) > 0,$$

as soon as the Poisson processes have non-zero intensities since for $j = 1, 2$, $dM_j^{[1]}(u) = \lambda_j(u)du$, with λ_j the intensity of X^j.

1.8.2 About the empirical centering assumption

Recall that

$$(\mathcal{A}^*_{Cent}) \quad \left| \begin{array}{l} \text{For } x_1 = (x_1^1, x_1^2), \ldots, x_n = (x_n^1, x_n^2) \text{ in } \mathcal{X}^2, \\ \sum_{i_1,i_2,i_1',i_2'=1}^n h\left(\left(x_{i_1}^1, x_{i_2}^2\right), \left(x_{i_1'}^1, x_{i_2'}^2\right)\right) = 0. \end{array} \right.$$

On the one hand, in the *Linear case*, that is if $h = h_\varphi$ (see (1.2.6)), then for $n \geq 1$ and for $x_1 = (x_1^1, x_1^2), \ldots, x_n = (x_n^1, x_n^2)$ in $\mathcal{X}^2$,

$$\begin{aligned}
&\sum_{i,i',j,j'=1}^{n} h\left(\left(x_i^1, x_{i'}^2\right), \left(x_j^1, x_{j'}^2\right)\right) \\
&\qquad = \frac{1}{2} \sum_{i,i',j,j'=1}^{n} \left(\varphi\left(x_i^1, x_{i'}^2\right) + \varphi\left(x_j^1, x_{j'}^2\right) - \varphi\left(x_i^1, x_{j'}^2\right) - \varphi\left(x_j^1, x_{i'}^2\right)\right) \\
&\qquad = \frac{n^2}{2}\left(\sum_{i,i'=1}^{n} \varphi\left(x_i^1, x_{i'}^2\right) + \sum_{j,j'=1}^{n} \varphi\left(x_j^1, x_{j'}^2\right) - \sum_{i,j'=1}^{n} \varphi\left(x_i^1, x_{j'}^2\right) - \sum_{j,i'=1}^{n} \varphi\left(x_j^1, x_{i'}^2\right)\right) \\
&\qquad = 0.
\end{aligned}$$

So $(\mathcal{A}^*_{Cent})$ is immediately satisfied in the *Linear case*.

On the other hand, $(\mathcal{A}^*_{Cent})$ does not imply that h is of the form h_φ.
Indeed, consider

$$h\left(\left(x^1, x^2\right), \left(y^1, y^2\right)\right) = \#x^1 \cdot \#x^2 \cdot \#y^1 \cdot \#y^2 \left[\left(\#x^1 - \#y^1\right)\left(\#x^2 - \#y^2\right)\right].$$

- The kernel h is obviously symmetric.
- The kernel h satisfies $(\mathcal{A}^*_{Cent})$. Indeed, let

  $$f\left(x^1, y^1\right) = \#x^1 \cdot \#y^1 \left(\#x^1 - \#y^1\right).$$

  First, notice that $f\left(x^1, x^1\right) = 0$ and $f\left(x^1, y^1\right) = -f\left(y^1, x^1\right)$.
  Moreover, $h\left(\left(x^1, x^2\right), \left(y^1, y^2\right)\right) = f\left(x^1, y^1\right) f\left(x^2, y^2\right)$. Thus

  $$\begin{aligned}
  &\sum_{i,i',j,j'=1}^{n} h\left(\left(x_i^1, x_{i'}^2\right), \left(x_j^1, x_{j'}^2\right)\right) \\
  &\qquad = \sum_{i,i',j,j'=1}^{n} f\left(x_i^1, x_j^1\right) f\left(x_{i'}^2, x_{j'}^2\right) \\
  &\qquad = \left(\sum_{i,j=1}^{n} f\left(x_i^1, x_j^1\right)\right)\left(\sum_{i',j'=1}^{n} f\left(x_{i'}^2, x_{j'}^2\right)\right) \\
  &\qquad = \left(\sum_{i=1}^{n} \underbrace{f\left(x_i^1, x_i^1\right)}_{0} + \sum_{1 \leq i < j \leq n} \underbrace{f\left(x_i^1, x_j^1\right) + f\left(x_j^1, x_i^1\right)}_{0}\right)\left(\sum_{i',j'=1}^{n} f\left(x_{i'}^2, x_{j'}^2\right)\right) \\
  &\qquad = 0,
  \end{aligned}$$

  and thus $(\mathcal{A}^*_{Cent})$ is satisfied by h.
- The kernel h cannot be written as an h_φ.
  On the one hand, first notice that for any $\varphi : \mathcal{X}^2 \to \mathbb{R}$, the difference

  $$D_{h_\varphi} := h_\varphi\left(\left(x^1, x^2\right), \left(y^1, y^2\right)\right) - h_\varphi\left(\left(\tilde{x}^1, x^2\right), \left(y^1, y^2\right)\right)$$

does not depend on y^1. Indeed,

$$\begin{aligned} D_{h_\varphi} &= \frac{1}{2}\Big(\varphi\left(x^1,x^2\right)+\varphi\left(y^1,y^2\right)-\varphi\left(x^1,y^2\right)-\varphi\left(y^1,x^2\right) \\ &\quad -\varphi\left(\tilde{x}^1,x^2\right)-\varphi\left(y^1,y^2\right)+\varphi\left(\tilde{x}^1,y^2\right)+\varphi\left(y^1,x^2\right)\Big) \\ &= \frac{1}{2}\left(\varphi\left(x^1,x^2\right)-\varphi\left(\tilde{x}^1,x^2\right)+\varphi\left(\tilde{x}^1,y^2\right)-\varphi\left(x^1,y^2\right)\right). \end{aligned}$$

On the other hand, for the kernel h introduced above, the difference D_h does depend on y^1. Indeed

$$\begin{aligned} D_h &= h\left(\left(x^1,x^2\right),\left(y^1,y^2\right)\right)-h\left(\left(\tilde{x}^1,x^2\right),\left(y^1,y^2\right)\right) \\ &= \#x^2\cdot\#y^1\cdot\#y^2\left(\#x^2-\#y^2\right)\times \\ &\quad \left[\#x^1\cdot\left(\#x^1-\#y^1\right)-\#\tilde{x}^1\cdot\left(\#\tilde{x}^1-\#y^1\right)\right], \end{aligned}$$

and if for instance, $\#x^1=\#y^2=1$ and $\#\tilde{x}^1=\#x^2=2$, then

$$D_h = 2\#y^1\left[-3+\#y^1\right],$$

which clearly depends on y^1.

So finally, there does not exist any φ such that $h=h_\varphi$.

Chapter 2

A distribution free UE method based on delayed coincidence count

In this chapter, we investigate several distribution free dependence detection procedures, mainly based on bootstrap and permutation principles and their approximation properties. Thanks to this study, we introduce a new distribution free Unitary Events (UE) method, named *Permutation UE method*, which consists in a multiple testing procedure based on permutation and delayed coincidence count. Each involved single test of this procedure achieves the prescribed level, so that the corresponding multiple testing procedure controls the False Discovery Rate (FDR), and this with as few assumptions as possible on the underneath distribution. Some simulations show that this method outperforms the trial-shuffling and the MTGAUE method in terms of single levels and FDR, for a comparable amount of false negatives. Application on real data is also provided.

This Chapter is the fruit of a collaboration with Yann Bouret[1], Magalie Fromont[2] and Patricia Reynaud-Bouret[3]. The corresponding article is currently in revision.

Contents

2.1 Introduction . . . 142
2.2 Binned and delayed coincidence counts . . . 143
2.3 Some distribution free independence tests . . . 146
2.3.1 A naive approach and the centering issue . . . 147
2.3.2 The bootstrap and permutation approaches . . . 150
2.3.3 Which centering for which resampling ? . . . 154
2.3.4 Practical testing procedures and p-values . . . 157
2.4 Multiple tests . . . 159
2.4.1 Description of the complete multiple testing algorithm . . . 159
2.4.2 Comparison on simulations . . . 160
2.4.3 Comparison on real data . . . 160
2.5 Conclusion . . . 162

[1]LPMC, Université Nice Sophia Antipolis
[2]IRMAR, Université Rennes 2
[3]LJAD, Université Nice Sophia Antipolis

2.1 Introduction

The eventual time dependence either between cerebral areas or between neurons, and in particular the synchrony phenomenon, has been vastly debated and investigated as a potential element of the neuronal code [166]. To detect such a phenomenon at the microscopic level, multielectrodes are usually used to record the nearby electrical activity. After pretreatment, the time occurrences of action potentials (spikes) for several neurons are therefore available. One of the first steps of analysis is then to understand whether and how two simultaneously recorded spike trains, corresponding to two different neurons, are dependent or not.
Several methods have been used to detect synchrony [3, 133]. Among the most popular ones, the UE method, due to Grün and collaborators [67, 71], has been applied in the last decade on a vast amount of real data (see, for instance, [109] and references therein). Two of its main features are at the root of its popularity: the UE method is not only able to give a precise location in time of the dependence periods, but also to quantify the degree of dependence by providing p-values for the independence tests.
From the initial method, substantial upgrades have been developed:

(i) In the original UE method, the point processes modeling the data are binned and clipped at a rough level, so that the final data have a quite low dimension (around a few hundreds per spike train). However, it is proved in [72] that the binned coincidence count, as a result of this pre-processing, may induce a loss in synchrony detection of about 60% in certain cases. The idea of [72] is therefore to keep the data at the initial resolution level despite its high dimension, and to define the notion of multiple shift (MS) coincidence count, nicely condensing the dependence feature that neurobiologists want to analyze without any loss in synchrony detection.

(ii) The original UE method assesses p-values by assuming that the underlying point processes are Poisson (or Bernoulli) processes. As there is still no commonly validated and accepted model for spike trains (see, for instance, [142] for a thorough data analysis), several surrogate data methods have been proposed [120]. In particular, trial-shuffling methods [137, 138] allow to assess p-values based on the assumption that i.i.d. trials are available, using bootstrap paradigm and without making any assumption on the underlying point processes distribution. However, up to our knowledge, surrogate data methods are always based on binned coincidence count (see section 2.2 for a precise definition) whose low complexity combined to parallel programming [44] make algorithms usable in practice.

(iii) The original UE method is based on two main statistical approximations. First, it involves the underlying intensities (or firing rates) of the Poisson processes, which are unknown in practice, and so replaced by estimates. However, this plug-in procedure is not taken into account in the p-value computations. Then, the detection of dependence through time is done by applying several tests at once without correcting for the multiplicity of the tests. In the recent work of [170], a multiple testing procedure based on a Gaussian approximation of the Unitary Events (MTGAUE) corrects those two facts, moreover including a generalization of the notion of MS coincidence count, namely the delayed coincidence count, which does not suffer from any loss in synchrony detection either. But MTGAUE is still based on the assumption that the underlying point processes are Poisson.

Our aim is here to go further by proposing a new delayed coincidence count-based multiple testing procedure, which does not need any binning pre-processing of the data as in [170], but which does not assume any model on the underlying point processes anymore. This procedure combines a permutation approach in the line of Hoeffding [84], Romano [155] and Romano and Wolf [156], with the multiple testing procedure of [16].
To do so, we first propose a fast algorithm to compute the delayed coincidence count, with a computational cost equivalent to the one of the binned coincidence count. Next we study several distribution free tests, most of them based on resampling approaches, as the trial-shuffling, or the permutation approach. We finally propose a complete multiple testing algorithm which satisfies similar properties as existing UE methods, but without sharing any of the previous drawbacks. Some simulations and applications to real data complete this study.

In all the sequel, X^1 and X^2 denote two point processes modeling the spike trains of two simultaneously recorded neurons and X represents the couple (X^1, X^2). The abbreviation "i.i.d." stands for independent and identically distributed. In this sense, by assuming that n independent trials are observed, the observation is modeled by an i.i.d. sample of size n of couples from the same distribution as X, meaning n i.i.d. copies $X_1, ..., X_n$ of X. This sample is denoted in the sequel by $\mathbb{X}_n = (X_1, ..., X_n)$. The corresponding probability and expectation are respectively denoted by $\mathbb{P}$ and $\mathbb{E}$.
Since the independence between X^1 and X^2 is the main focus of the present work, the following notation is useful: $X^{\perp\!\!\!\perp}$ denotes a couple $(X^{1,\perp\!\!\!\perp}, X^{2,\perp\!\!\!\perp})$ such that $X^{1,\perp\!\!\!\perp}$ (resp. $X^{2,\perp\!\!\!\perp}$) has the same distribution as X^1 (resp. X^2), but $X^{1,\perp\!\!\!\perp}$ is independent of $X^{2,\perp\!\!\!\perp}$. In particular, the couple $X^{\perp\!\!\!\perp}$ has the same marginals as the couple X. Moreover, $\mathbb{X}_n^{\perp\!\!\!\perp}$ denotes an i.i.d. sample of size n from the distribution of $X^{\perp\!\!\!\perp}$, and $\mathbb{P}_{\perp\!\!\!\perp}$ and $\mathbb{E}_{\perp\!\!\!\perp}$ are the corresponding probability and expectation. Note in particular that if the two observed neurons indeed behave independently, then the observed sample $\mathbb{X}_n$ has the same distribution as $\mathbb{X}_n^{\perp\!\!\!\perp}$.
The notation $\mathbf{1}_A$ stands for a function whose value is 1 if the event A holds and 0 otherwise. Finally, for any point process X^j $(j = 1, 2)$, dN_{X^j} stands for its associated point measure, defined for all measurable function f by

$$\int f(u) dN_{X^j}(u) = \sum_{T \in X^j} f(T),$$

and for any interval I, $N_{X^j}(I)$ denotes the number of points of X^j observed in I.

2.2 Binned and delayed coincidence counts

Because of the way neurons transmit information through action potentials, it is commonly admitted that the dependence between the spike trains of two neurons is due to temporal correlations between spikes produced by both neurons. Informally, a coincidence occurs when two spikes (one from each neuron) appear with a delay less than a fixed δ (of the order of a few milliseconds). Several coincidence count functions have been defined in the neuroscience literature, and among them the classical binned coincidence count, introduced in [69, 70].

Definition 2.2.1. The binned coincidence count between point processes X^1 and X^2 on the interval $[a, b]$ with $b - a = M\delta$ for an integer $M \geq 2$ and a fixed delay $\delta > 0$ is given by

$$\psi_\delta^{coinc}(X^1, X^2) = \sum_{\ell=1}^{M} \mathbf{1}_{N_{X^1}(I_\ell)\geq 1} \mathbf{1}_{N_{X^2}(I_\ell)\geq 1},$$

where I_ℓ is the ℓth bin of length δ, that is $[a+(\ell-1)\delta, a+\ell\delta)$, and

$$\mathbf{1}_{N_{X^j}(I_\ell)\geq 1} = \begin{cases} 1 & \text{if there is at least one point of } X^j \text{ in the } \ell\text{th bin,} \\ 0 & \text{if there is no point of } X^j \text{ in the } \ell\text{th bin.} \end{cases}$$

More informally, the binned coincidence count is the number of bins that contain at least one spike of each spike trains, as one can see on Figure 2.1.
The binned coincidence count computation algorithm is usually performed on already binned and clipped data (see Figure 2.1). Therefore, given two sequences of 0 and 1 of length $(b-a)\,\delta^{-1}$, $2(b-a)\delta^{-1}$ operations are needed to compute the binned coincidence count, without counting the number of operations needed for the binning pre-processing.
The more recent notion of delayed coincidence count, introduced in [170], is a generalization of the multiple-shift coincidence count, defined in [72] for discretized point processes, to not necessarily discretized point processes.

Definition 2.2.2. The delayed coincidence count between point processes X^1 and X^2 on the interval $[a,b]$ is given by

$$\varphi_\delta^{coinc}(X^1, X^2) = \int_a^b \int_a^b \mathbf{1}_{|u-v|\leq\delta} dN_{X^1}(u) dN_{X^2}(v),$$

More informally, $\varphi_\delta^{coinc}(X^1, X^2)$ is the number of couples of spikes (one spike from X^1 and one from X^2) appearing in $[a,b]$ with delay at most equal to δ. A visual example is given on Figure 2.1. Note in particular that two coincidences are discarded by the binned coincidence count on this particular example: one because if the clipping effect in the third bin and one because of the effect of adjacent bins in the seventh and eighth bins. Both of them are counted in the delayed coincidence count.
However, a rather naive algorithm would test whether for any pair (u,v) of a spike u in X^1 and a spike v in X^2, the delay $|u-v|$ is less than δ and to count the number of hits. This would lead to an algorithm whose complexity is in the product of the number of points in each spike trains. If one assumes both spike trains to be Poisson with intensity λ_1 and λ_2, this algorithm costs in average $\lambda_1\lambda_2(b-a)^2$ and is therefore quadratic in the length of the interval. One can actually drastically improve this rate thanks to the following algorithm for which the result $c := \varphi_\delta^{coinc}(X^1, X^2)$ is the delayed coincidence count.

Delayed coincidence count algorithm

Given two sequences x_1 and x_2 of ordered points with respective lengths $n_1 = N_{X^1}([a,b])$ and $n_2 = N_{X^2}([a,b])$, representing the observations of two point processes X^1 and X^2 on the interval $[a,b]$,

- Initialize $j=1$ and $c=0$.
- For $i = 1, ..., n_1$,
  1. Assign $x_{low} = x_1[i] - \delta$.
  2. While $j \leq n_2$ and $x_2[j] < x_{low}$, $j = j+1$.
  3. If $j > n_2$, stop.
  4. Else (here necessarily, $x_2[j] \geq x_{low}$),
     4.a Assign $x_{up} = x_1[i] + \delta$ and $k = j$.
     4.b While $k \leq n_2$ and $x_2[k] \leq x_{up}$, $c = c+1$ and $k = k+1$.

This algorithm is slightly more involved but the computational complexity is much smaller than the previous one. Figure 2.1 gives a visualisation of the algorithm on a very simple

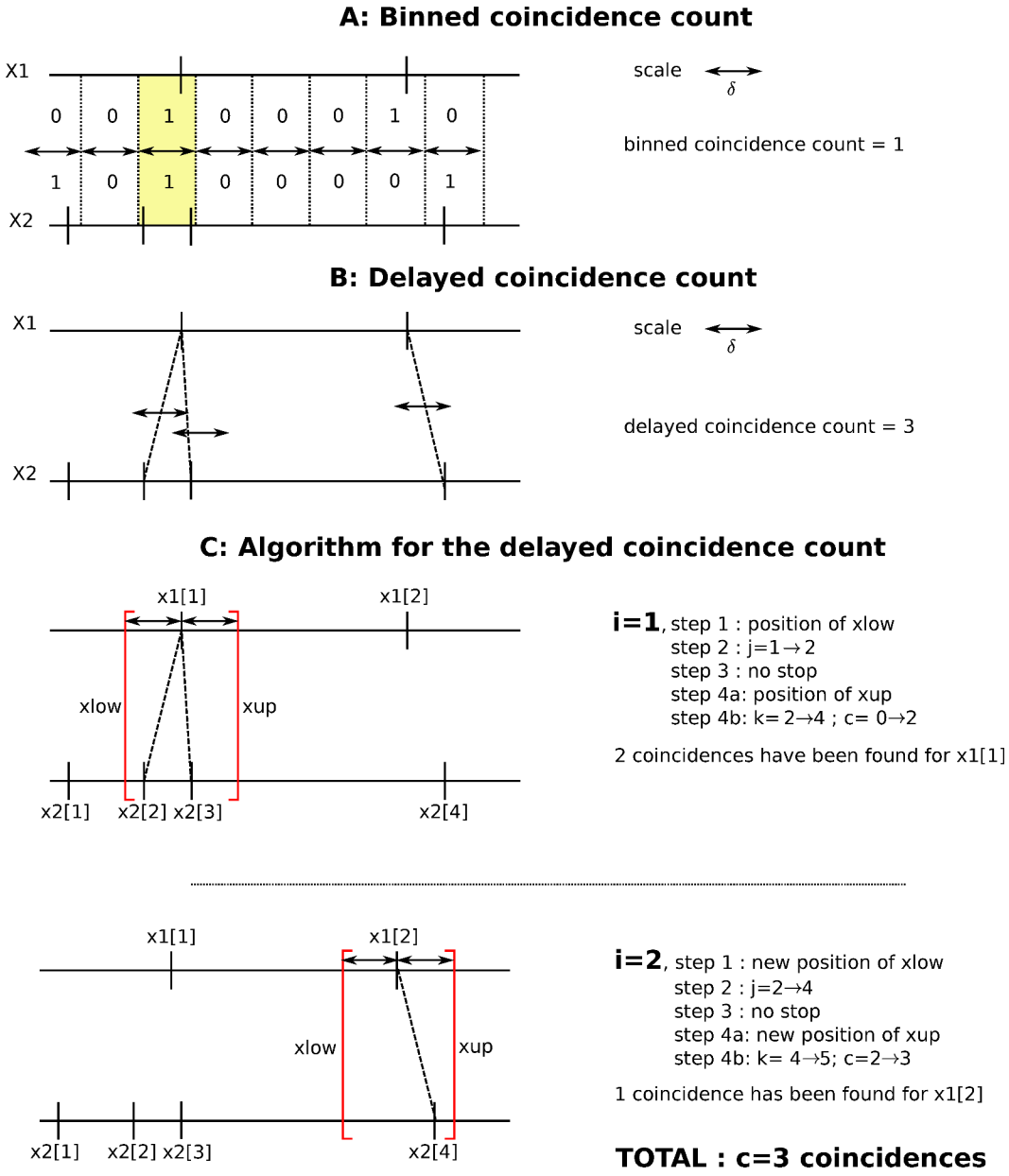


Figure 2.1 – Coincidence counts. Part **A** gives an example of binned coincidence count on a couple of spike trains (X^1, X^2) (the spikes corresponding to the respective dashes on the line): after binning the data into blocks of length δ, one only keeps the information whether there is at least a spike or not in the bin (clipping). The binned coincidence count is then the number of times there is a "1" for each spike train in the same bin. Part **B** gives, on the same example, the number of delayed coincidence count, that is the number of pairs of points (one on each spike train) at distance less than δ. Note that those two coincidence count are on this particular example different. Part **C** visualizes what the proposed algorithm is doing. In particular, note that it exploits the sparsity of the data represented via the vector x_1 and x_2: there is no computational time spent on the central part of the drawing corresponding to the 0's of the binned process.

example. The main point is that the index j cannot decrease and therefore it is not making a double full loop on all the indices of both sequences x_1 and x_2. A pseudo double loop is made thanks to the index k which indeed can take several times the same value but whose range is only governed by the number of points that appear in an interval of length 2δ, namely $[x_{low}, x_{up}]$, which is usually much smaller than the total length of the sequence x_2. More precisely, the complexity of the algorithm is therefore bounded by $3n_1$ (for steps 1, 3 and 4.a), plus n_2 (for all steps 2 on all points of x_1, that is the range of the index j which never decreases), plus roughly $2n_1$ times the number of points of x_2 in a segment (namely $[x_{low}, x_{up}]$) of length 2δ (for step 4.b). In average, if X^1 and X^2 are for instance independent homogeneous Poisson processes of respective intensities λ_1 and λ_2, at most $3\lambda_1(b-a)+\lambda_2(b-a)+2\lambda_1(b-a)(2\delta\lambda_2)$ operations are required. So, for typical parameters ($(b-a) = 0.1$s, $\delta = 0.005$s, $\lambda_1 = \lambda_2 =$ 50Hz), 40 operations in average are required to compute the binned coincidence count, against 25 operations for the delayed coincidence count. Both algorithms are therefore linear in $(b-a)$ with a slight advantage for the delayed coincidence count which exploits the sparsity of the spike trains, in the usual range of parameters (see also Figure 2.1 for a more visual representation of this sparsity, the bins with 0's being not even coded in the present algorithm). Notice that all surrogate data methods (see [120]) could in principle be applied on this new notion of coincidence, at least when only two neurons are involved.

2.3 Some distribution free independence tests

Given the observation of a n sample $\mathbb{X}_n = (X_1, \ldots, X_n)$ corresponding to n different trials, the aim is to test:

$$(\mathcal{H}_0) \text{ "}X^1 \text{ and } X^2 \text{ are independent on } [a,b]\text{"}$$

against

$$(\mathcal{H}_1) \text{ "}X^1 \text{ and } X^2 \text{ are not independent on } [a,b]\text{".}$$

All existing UE methods are based on a statistic equal to the total number of coincidences

$$\mathbf{C} = \mathbf{C}(\mathbb{X}_n) = \sum_{i=1}^{n} \varphi(X_i^1, X_i^2),$$

where φ generically denotes either φ_δ^{coinc}, or ψ_δ^{coinc}, or other coincidence count functions that practitioners would like to use (see Chapter 1 for other choices).

To underline what is observed or not, when $\mathbf{C}$ is computed on the observation of $\mathbb{X}_n$, it is denoted by $\mathbf{C}^{obs}$, representing then the total number of observed coincidences.

In the following, several of these UE methods or testing procedures are described, which all rely on the same paradigm: "reject $(\mathcal{H}_0)$ when $\mathbf{C}^{obs}$ is significantly different from what is expected under $(\mathcal{H}_0)$". More precisely, the independence $(\mathcal{H}_0)$ is rejected and the dependence is detected when a quantity, based on the difference between the observed coincidence count and what is expected under $(\mathcal{H}_0)$, is smaller or larger than some critical values. Those critical values are obtained in various ways, each of them being peculiar to each method. Note that the following procedures could be applied to any chosen coincidence count function φ, though the implemented procedures of the present simulations and data analysis only use the delayed coincidence count, that is $\varphi = \varphi_\delta^{coinc}$.

2.3.1 A naive approach and the centering issue

As noticed above, the only question, when considering UE methods, is how to construct the critical values.
If the values of the expectation and the variance of $\mathbf{C}$ under $(\mathcal{H}_0)$, that is

$$c_0 = \mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}] \text{ and } v_0 = \mathbb{E}_{\perp\!\!\!\perp}\left[(\mathbf{C} - c_0)^2\right],$$

are precisely known, then the classical central limit theorem gives under independence that

$$\frac{\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp}) - c_0}{\sqrt{v_0}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Then, given α in $(0,1)$, the test which consists in rejecting $(\mathcal{H}_0)$ when $(\mathbf{C}^{obs} - c_0)/\sqrt{v_0}$ is larger than $z_{1-\alpha}$, the $1-\alpha$ quantile of a standard Gaussian distribution, is asymptotically of level α. It means that, for this test, the probability of rejecting independence, whereas independence holds, is asymptotically (in n, the number of trials) smaller than the prescribed α.
In the present point processes framework, strong distribution assumptions, for which the values of c_0 and v_0 are precisely known, are unrealistic. Even if the spike trains are assumed to be homogeneous Poisson processes as in [170], those quantities depend, through some formulas, on the unknown firing rates that have to be estimated and plugged into these precise formulas. It has been shown that this modifies the asymptotic variance shape, and tests under Poisson assumptions with unknown firing rates have been developed in [170].
Since Poisson assumptions are quite doubtful on real data [10, 53, 147], the aim of the present work is to go further by not assuming any Poisson or other model assumptions for the spike trains. In this sense, the aim is to develop "distribution free" methods that are completely agnostic with respect to the underlying distribution of the spike trains. In this case, a preliminary step is to estimate c_0, only using the sample $\mathbb{X}_n$ without any distribution assumption. Note that

$$c_0 = \mathbb{E}_{\perp\!\!\!\perp}\left[\sum_{i=1}^{n} \varphi(X_i^{1,\perp\!\!\!\perp}, X_i^{2,\perp\!\!\!\perp})\right] = n\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi(X^{1,\perp\!\!\!\perp}, X^{2,\perp\!\!\!\perp})\right],$$

and that for $i \neq i'$, as X_i is always assumed to be independent of $X_{i'}$,

$$\mathbb{E}\left[\varphi(X_i^1, X_{i'}^2)\right] = \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi(X^{1,\perp\!\!\!\perp}, X^{2,\perp\!\!\!\perp})\right]. \tag{2.3.1}$$

Therefore, c_0 can always be estimated in a distribution free manner by

$$\hat{\mathbf{C}}_0(\mathbb{X}_n) = \frac{1}{n-1}\sum_{i\neq i'} \varphi(X_i^1, X_{i'}^2).$$

Hence a reasonable test statistic would be based on the difference:

$$\mathbf{U} = \mathbf{U}(\mathbb{X}_n) = \mathbf{C}(\mathbb{X}_n) - \hat{\mathbf{C}}_0(\mathbb{X}_n),$$

its observed version being denoted by $\mathbf{U}^{obs}$. Here, $\mathbf{U}(\mathbb{X}_n)$ is not an empirical mean, but a U-statistic, so it does not satisfy the classical central limit theorem. Hence, its limit distribution under $(\mathcal{H}_0)$ is not as straightforward as usual. Nevertheless, some asymptotic theorems, proved in Chapter 1, Section 1.3.4 (see Proposition 1.3.5 with the equivalence

$\mathbf{U}(\mathbb{X}_n) = n \times U_{n,h_\varphi}(\mathbb{X}_n))$, show that under mild conditions (always satisfied in practice in the present cases) the following convergence result holds

$$\mathbf{Z}(\mathbb{X}_n^{\perp\!\!\!\perp}) = \frac{\mathbf{U}(\mathbb{X}_n^{\perp\!\!\!\perp})}{\sqrt{n}\hat{\sigma}(\mathbb{X}_n^{\perp\!\!\!\perp})} \underset{n\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}(0,1), \tag{2.3.2}$$

where

$$\hat{\sigma}^2(\mathbb{X}_n) = \frac{4}{n(n-1)(n-2)} \sum_{i,j,k \text{ all different}} h(X_i, X_j) h(X_i, X_k),$$

with

$$h(x,y) = \frac{1}{2}\Big[\varphi(x^1,x^2) + \varphi(y^1,y^2) - \varphi(x^1,y^2) - \varphi(y^1,x^2)\Big].$$

As above, denoting by $\mathbf{Z}^{obs}$ the quantity $\mathbf{Z}$ computed on the observed sample, (2.3.2) implies that for some fixed α in $(0,1)$, the test that consists in rejecting $(\mathcal{H}_0)$ when $\mathbf{Z}^{obs} \geq z_{1-\alpha}$, is asymptotically of level α.
The approximation properties of (2.3.2) are illustrated on Figure 2.2.

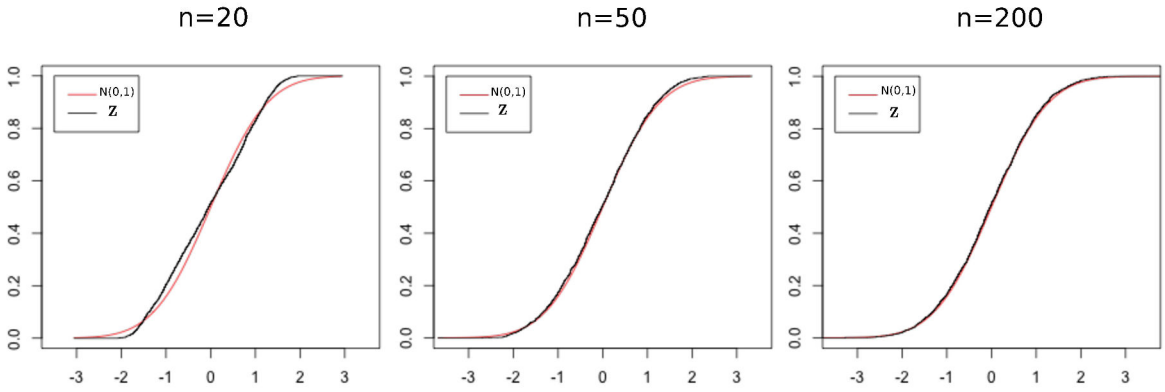


Figure 2.2 – Gaussian approximation of the distribution of $\mathbf{Z}$. The black line corresponds to the cumulative distribution function (c.d.f.) of $\mathbf{Z}$ under $(\mathcal{H}_0)$, that is of $\mathbf{Z}^{\perp\!\!\!\perp} = \mathbf{Z}(\mathbb{X}_n^{\perp\!\!\!\perp})$, obtained with 2,000 simulations of $\mathbb{X}_n^{\perp\!\!\!\perp}$, for $n = 20$, 50 or 200 trials of two independent Poisson processes of firing rate 30Hz, on a window of length 0.1s with $\delta = 0.01$s. The red line corresponds to the standard Gaussian c.d.f.

Clearly, one can see that the distribution approximation is good when n is large ($n = 200$) as expected, but not so convincing for small values of n ($n = 20$, or even $n = 50$), particularly in the tail parts of the distributions. However, as it is especially the tails of the distributions that are involved in the test through the quantile $z_{1-\alpha}$, one can wonder, by looking at Figure 2.2, if it may perform reasonably well in practice with a usual number of a few tens of trials. Furthermore, looking informally at equation (2.3.2), readers could think of two approximations that could be roughly formulated in the following way:

$$\mathbf{U}(\mathbb{X}_n^{\perp\!\!\!\perp}) \underset{n\to\infty}{\overset{\mathcal{L}}{\approx}} \mathcal{N}\left(0, n\hat{\sigma}^2(\mathbb{X}_n^{\perp\!\!\!\perp})\right), \tag{2.3.3}$$

and

$$\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp}) \underset{n\to\infty}{\overset{\mathcal{L}}{\approx}} \mathcal{N}\left(\hat{\mathbf{C}}_0(\mathbb{X}_n^{\perp\!\!\!\perp}), n\hat{\sigma}^2(\mathbb{X}_n^{\perp\!\!\!\perp})\right). \tag{2.3.4}$$

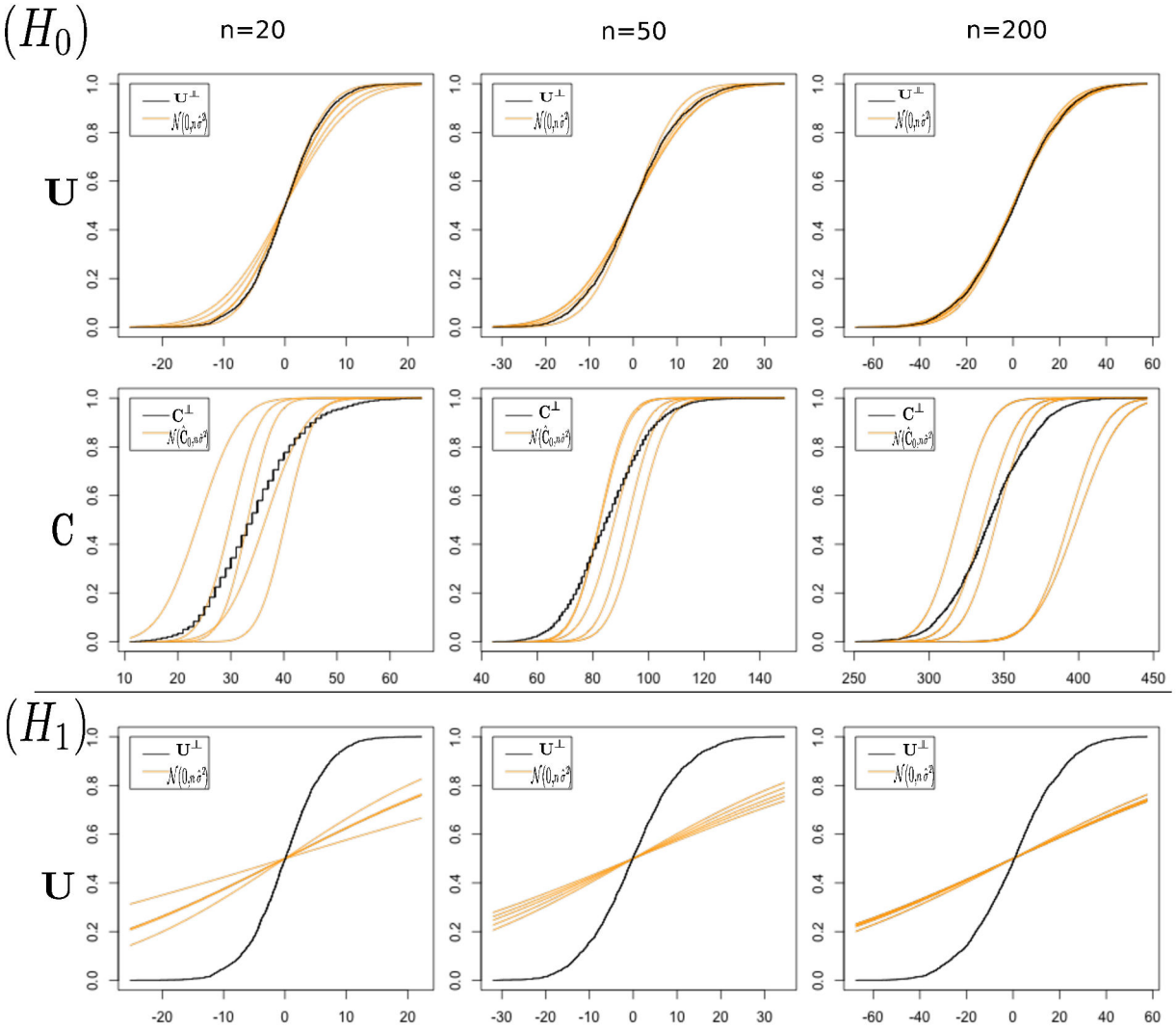


Figure 2.3 – Other Gaussian distribution approximations. Two first lines: in black, c.d.f. of $\mathbf{U}$ and $\mathbf{C}$ under $(\mathcal{H}_0)$, obtained as in Figure 2.2. These c.d.f. are respectively compared with the Gaussian c.d.f. with mean 0 and standard deviation $\sqrt{n}\hat{\sigma}(\mathbb{X}_n)$ (first line), and the Gaussian c.d.f. with mean $\hat{\mathbf{C}}_0(\mathbb{X}_n)$ and standard deviation $\sqrt{n}\hat{\sigma}(\mathbb{X}_n)$ (second line), for five different simulations of $\mathbb{X}_n$ under $(\mathcal{H}_0)$, all plotted in orange. Third line: in black, c.d.f. of $\mathbf{U}$ under $(\mathcal{H}_0)$ computed as above, compared with the centered Gaussian c.d.f. with standard deviation $\sqrt{n}\hat{\sigma}(\mathbb{X}_n)$ (in orange), for five different simulations of $\mathbb{X}_n$ under $(\mathcal{H}_1)$ (same marginals as in the first two lines but $X^1 = X^2$).

This is illustrated on Figure 2.3.

Looking at the first line of Figure 2.3, one can see that the approximation formulated in (2.3.3) is actually conceivable for large values of n. Note that in practice, one cannot have access to $\hat{\sigma}^2(\mathbb{X}_n^{\perp\!\!\!\perp})$ and it has to be replaced by $\hat{\sigma}^2(\mathbb{X}_n)$, meaning that it is computed with the observed sample. This does not change anything under $(\mathcal{H}_0)$ since $\mathbb{X}_n$ is, in this case, distributed as $\mathbb{X}_n^{\perp\!\!\!\perp}$. But this is a particularly important sticking point if $(\mathcal{H}_0)$ is not satisfied, as one can see on the third line of Figure 2.3: the distribution of $\mathbf{U}(\mathbb{X}_n^{\perp\!\!\!\perp})$ does not look like a centered Gaussian distribution of variance $n\hat{\sigma}^2(\mathbb{X}_n)$, when $\mathbb{X}_n$ does not satisfy $(\mathcal{H}_0)$.

More importantly, the second line of Figure 2.3 shows that the approximation formulated in (2.3.4) is in fact misleading. To understand why, one needs to take into account the two following points.

(i) *On the mean.* $\hat{\mathbf{C}}_0(\mathbb{X}_n^{\perp\!\!\!\perp})$ moves around its expectation c_0 (which is also the expectation of $\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp})$) with realizations of $\mathbb{X}_n^{\perp\!\!\!\perp}$. These fluctuations have an order of magnitude of $\sqrt{n}$ and are therefore perfectly observable on the distribution of $\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp})$ whose variance is also of order $\sqrt{n}$.

(ii) *On the variance.* $n\hat{\sigma}^2(\mathbb{X}_n^{\perp\!\!\!\perp})$ estimates the variance of $\mathbf{U}(\mathbb{X}_n^{\perp\!\!\!\perp})$ and not the one of $\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp})$ or $\hat{\mathbf{C}}_0(\mathbb{X}_n^{\perp\!\!\!\perp})$. This explains why not only the mean but also the variance are badly estimated in the second line of Figure 2.3. Both randomness (the one of $\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp})$ and the one of $\hat{\mathbf{C}}_0(\mathbb{X}_n^{\perp\!\!\!\perp})$) have to be taken into account to estimate the variance of $\mathbf{U}(\mathbb{X}_n^{\perp\!\!\!\perp})$.

As a conclusion of this first naive approach, the test of purely asymptotic nature, which consists in rejecting $(\mathcal{H}_0)$ when $\mathbf{Z}^{obs} > z_{1-\alpha}$ may work for n large enough, as the variance is here computed by considering the correctly recentered statistic $\mathbf{U}$, and this even if the behavior of the statistic under $(\mathcal{H}_1)$ is not clear. But an ad hoc and more naive test statistic, based on an estimation of the variance of $\mathbf{C}$ directly and without taking into account the fact that the centering term $\hat{\mathbf{C}}_0(\mathbb{X}_n)$ is also random, would not lead to a meaningful test.

2.3.2 The bootstrap and permutation approaches

It is well known [62] that tests of purely asymptotic nature as the one presented above are less accurate for small values of n than more involved procedures. In this chapter, the focus is on bootstrap type procedures that are usually known to improve the performance from moderate to large sample sizes. Three main procedures are investigated: the trial-shuffling introduced in [138, 137], the classical full bootstrap test of independence and the permutation approach as in [155] or Chapter 1.

The main common paradigm of these three methods, as described in the sequel, is that starting from an observation of the sample $\mathbb{X}_n$, they randomly generate via a computer another sample $\tilde{\mathbb{X}}_n$, whose distribution should be close to the distribution of $\mathbb{X}_n^{\perp\!\!\!\perp}$ (see also Figure 2.5).

Trial-shuffling

$$\tilde{\mathbb{X}}_n = \mathbb{X}_n^{TS} = \left(\left(X^1_{i^{TS}(1)}, X^2_{j^{TS}(1)} \right), ..., \left(X^1_{i^{TS}(n)}, X^2_{j^{TS}(n)} \right) \right),$$

where the $(i^{TS}(k), j^{TS}(k))$'s are n i.i.d. couples drawn uniformly at random in $\{(i,j) \ / \ i = 1, ..., n, j = 1, ..., n, i \neq j\}$.
In particular, the corresponding bootstrapped coincidence count is

$$\mathbf{C}^{TS} = \mathbf{C}(\mathbb{X}_n^{TS}) := \sum_{k=1}^{n} \varphi\Big(X^1_{i^{TS}(k)}, X^2_{j^{TS}(k)} \Big) .$$

This algorithm seems natural with respect to (2.3.1) because it avoids the diagonal terms of the square $\{(i,j) \ / \ i = 1, ..., n, \ j = 1, ..., n\}$. Hence as a result,

$$\mathbb{E}\left[\mathbf{C}^{TS}\right] = c_0 = \mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}] .$$

Classical full bootstrap

$$\tilde{\mathbb{X}}_n = \mathbb{X}_n^{*} = \left(\left(X^1_{i^*(1)}, X^2_{j^*(1)} \right), ..., \left(X^1_{i^*(n)}, X^2_{j^*(n)} \right) \right),$$

where the n couples $(i^*(k), j^*(k))$ are i.i.d. and where $i^*(k)$ and $j^*(k)$ are drawn uniformly and independently at random in $\{1, ..., n\}$.
In particular, the corresponding bootstrapped coincidence count is

$$\mathbf{C}^{*} = \mathbf{C}(\mathbb{X}_n^{*}) := \sum_{k=1}^{n} \varphi\Big(X^1_{i^*(k)}, X^2_{j^*(k)} \Big) .$$

Note that this algorithm draws uniformly at random in the square $\{(i,j) \ / \ i = 1, ..., n, \ j = 1, ..., n\}$ and therefore does not avoid the diagonal terms. The idea behind this algorithm is to mimic the independence under $(\mathcal{H}_0)$ of X_k^1 and X_k^2 by drawing the indexes $i^*(k)$ and $j^*(k)$ independently. However,

$$\mathbb{E}\left[\mathbf{C}^{*}\right] = n \left(\frac{1}{n} \mathbb{E}\left[\varphi(X^1, X^2)\right] + \frac{n-1}{n} \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi(X^{1,\perp\!\!\!\perp}, X^{2,\perp\!\!\!\perp})\right] \right) .$$

Hence under $(\mathcal{H}_0)$, $\mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}^{*}] = c_0$ but, under $(\mathcal{H}_1)$, $\mathbb{E}\left[\mathbf{C}^{*}\right]$ and c_0 are only asymptotically equivalent.

Permutation

$$\tilde{\mathbb{X}}_n = \mathbb{X}_n^{\Pi_n} = \left(\left(X^1_1, X^2_{\Pi_n(1)} \right), ..., \left(X^1_n, X^2_{\Pi_n(n)} \right) \right),$$

where Π_n is a permutation drawn uniformly at random in the group of permutations $\mathfrak{S}_n$ of the set of indexes $\{1, \ldots, n\}$.
In particular, the corresponding permuted coincidence count is

$$\mathbf{C}^{\star} = \mathbf{C}\left(\mathbb{X}_n^{\Pi_n}\right) := \sum_{i=1}^{n} \varphi\Big(X^1_i, X^2_{\Pi_n(i)} \Big) .$$

The idea is to use permutations to avoid picking twice the same spike train of the same trial. In particular under $(\mathcal{H}_0)$, the sum in $\mathbf{C}^\star$ is still a sum of independent variables, which is not the case in both of the previous algorithms. However, under $(\mathcal{H}_1)$, the behavior is not as limpid. As for the full bootstrap,

$$\mathbb{E}\left[\mathbf{C}^\star\right] = n\left(\frac{1}{n}\mathbb{E}\left[\varphi(X^1, X^2)\right] + \frac{n-1}{n}\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi(X^{1,\perp\!\!\!\perp}, X^{2,\perp\!\!\!\perp})\right]\right).$$

Hence, under $(\mathcal{H}_1)$, $\mathbb{E}\left[\mathbf{C}^*\right]$ and c_0 are only asymptotically equivalent.

To compare those three resampling algorithms, the first thing to wonder is whether, at least under $(\mathcal{H}_0)$, the introduced extra randomness has not impacted the distribution. More precisely, as stated above, all three procedures satisfy

$$\mathbb{E}_{\perp\!\!\!\perp}\left[\mathbf{C}(\tilde{\mathbb{X}}_n)\right] = \mathbb{E}_{\perp\!\!\!\perp}\left[\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp})\right] = c_0,$$

but is the full distribution of $\mathbf{C}(\tilde{\mathbb{X}}_n)$ the same as the one of $\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp})$?
The first line of Figure 2.4 shows as expected that the permutation does not change the distribution of $\mathbb{X}_n^{\perp\!\!\!\perp}$, since, as said above, no spike train is picked twice. However, clearly the trial-shuffling and the full bootstrap have not the same property, even if the distributions are quite close.
Nevertheless, this is not completely convincing. Indeed, the main particularity of resampling procedures, such as bootstrap or permutation, is to be able for one current observation of $\mathbb{X}_n$ to generate several realizations of $\tilde{\mathbb{X}}_n$ to obtain not the unconditional distribution of $\mathbf{C}(\tilde{\mathbb{X}}_n)$ but the conditional distribution of $\mathbf{C}(\tilde{\mathbb{X}}_n)$ given $\mathbb{X}_n$. Figure 2.5 gives a more visual representation of the difference between conditional and unconditional distributions.
In particular, in a bootstrap (respectively permutation) testing procedure, the critical values are quantiles [4] of the conditional distribution of the bootstrapped (respectively permuted) test statistic given the observation of $\mathbb{X}_n$ and not the quantiles of the unconditional distribution. Hence, to see whether bootstrapped or permuted critical values associated to $\mathbf{C}(\mathbb{X}_n)$ are reasonable, the conditional distribution of $\mathbf{C}(\tilde{\mathbb{X}}_n)$ given $\mathbb{X}_n$ has to be compared with the distribution of $\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp})$, and this whether $\mathbb{X}_n$ satisfies $(\mathcal{H}_0)$ or not.
The second line of Figure 2.4 shows what happens for the approximated conditional distribution of $\mathbf{C}(\tilde{\mathbb{X}}_n)$ given $\mathbb{X}_n$ under $(\mathcal{H}_0)$ in the three considered resampling approaches. Surprisingly none of these three conditional distributions seems to fit the distribution of $\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp})$. One may eventually think that this is due to the Monte Carlo approximation of the conditional distributions, but for the trial-shuffling approach, Pipa and Grün developed an algorithm for exact computation of the conditional distribution [137]: both Monte Carlo and exact conditional distribution are so close that it is difficult to make any difference between them. Hence there should be another explanation. In fact, the curves on the second line of Figure 2.4 are similar to the ones on the second line of Figure 2.3. In both set-ups, one wonders if the distribution of $\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp})$ can or cannot be approximated by a distribution depending on the observation of $\mathbb{X}_n$: a very basic Gaussian distribution for Figure 2.3 and a more intricate distribution using the bootstrap paradigm for Figure 2.4. Nevertheless both distributions are too widely spread around the aim which is the distribution of $\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp})$. Since the explanation for Figure 2.3 was a centering defect that can be corrected by considering $\mathbf{U}$, one of the possible explanation here is a centering defect for the bootstrap or permutation procedures too.

[4] In fact, the quantiles are usually approximated by a Monte Carlo method, since, in practice, one has access to a huge, but still reasonable number of realizations of $\tilde{\mathbb{X}}_n$ given $\mathbb{X}_n$.

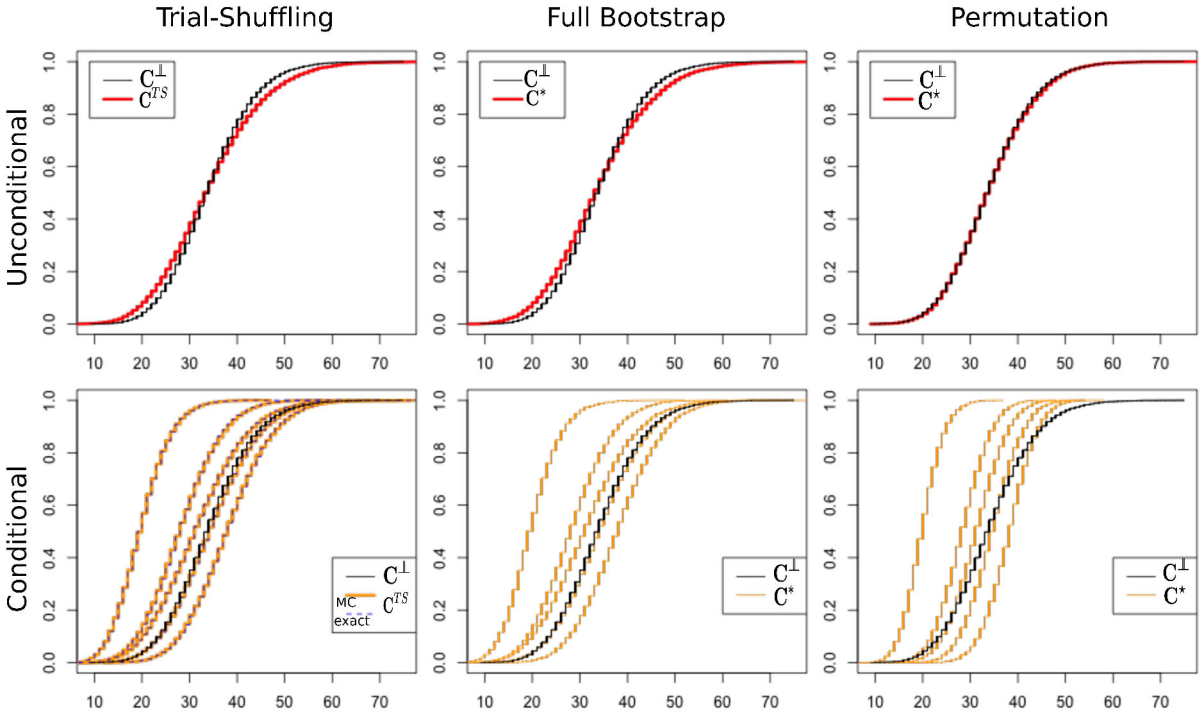


Figure 2.4 – The unconditional distribution and conditional distributions of $\mathbf{C}$ under $(\mathcal{H}_0)$. C.d.f. of $\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp})$ in black, and (for the first line) of $\mathbf{C}^{TS} = \mathbf{C}(\mathbb{X}_n^{TS})$, of $\mathbf{C}^* = \mathbf{C}(\mathbb{X}_n^*)$ and of $\mathbf{C}^\star = \mathbf{C}(\mathbb{X}_n^{\Pi_n})$ in red obtained from 10,000 simulations of $n = 20$ trials of two independent Poisson processes of firing rate 30Hz on a window of length 0.1s with $\delta = 0.01$s. On the second line, in addition to the c.d.f. of $\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp})$, five observations of $\mathbb{X}_n = \mathbb{X}_n^{\perp\!\!\!\perp}$ have been simulated in the same set-up and given these observations, the conditional c.d.f. of $\mathbf{C}^{TS}$, of $\mathbf{C}^*$ and of $\mathbf{C}^\star$ (in orange) have been approximated by simulating 10,000 times the extra-randomness corresponding to $\tilde{\mathbb{X}}_n$. For the trial-shuffling, in addition to this approximate Monte Carlo method (MC), the exact conditional c.d.f. (in dashed grey) has been obtained thanks to the algorithm of [137].

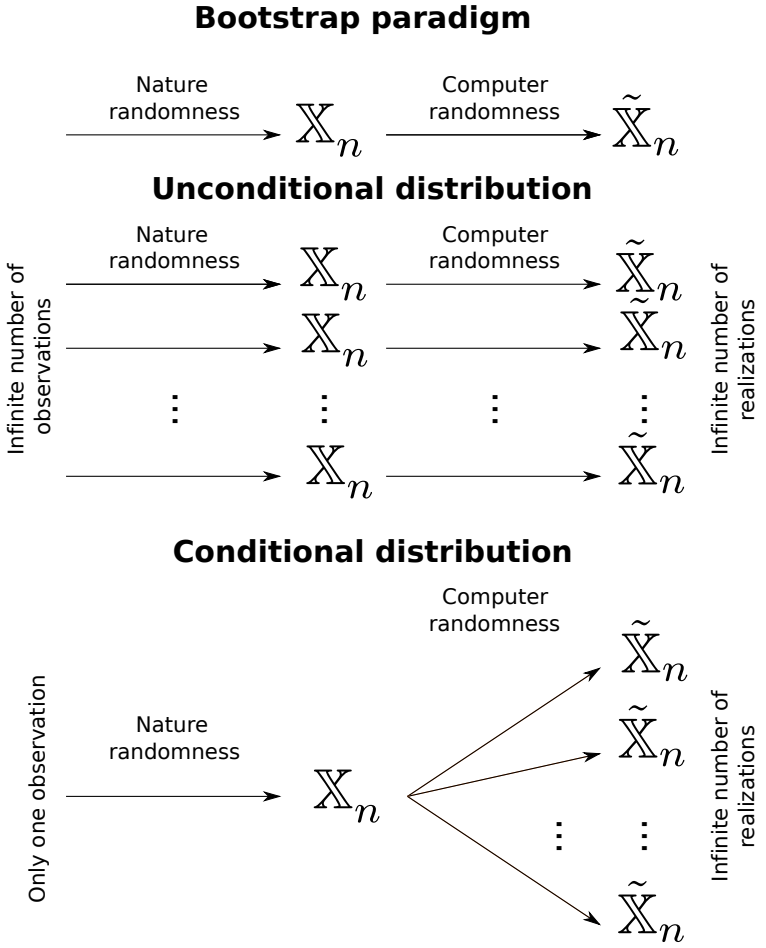


Figure 2.5 – Schematic view of the bootstrap paradigm and the difference between unconditional distribution and conditional distribution given the observation.

2.3.3 Which centering for which resampling ?

All the bootstrap approaches that have been proved to work from a mathematical point of view are based on centered quantities [62]. In particular, the precursor work of Bickel and Freedman [20] on the bootstrap of the mean can be heuristically explained as follows.
Given a n sample of i.i.d. real random variables $\mathbb{Y}_n = (Y_1, ..., Y_n)$ with mean m and a corresponding bootstrap sample $\mathbb{Y}_n^*$, it is not possible to estimate the distribution of the empirical mean $\bar{Y} = (1/n)\sum_{i=1}^n Y_i$ directly. However one can estimate the centered distribution, that is the distribution of $\bar{Y} - m = \bar{Y} - \mathbb{E}\left[\bar{Y}\right]$. To do so, it is sufficient to replace "empirical mean" by "empirical bootstrap mean" and "expectation" by "conditional expectation". More explicitly, denoting by $\bar{Y}^*$ the empirical mean of the bootstrap sample $\mathbb{Y}_n^*$, the distribution of $\bar{Y} - \mathbb{E}\left[\bar{Y}\right]$ is approximated by the conditional distribution of $\bar{Y}^* - \mathbb{E}\left[\bar{Y}^* | \mathbb{Y}_n\right]$ given $\mathbb{Y}_n$.

Transposed in our framework, this paradigm would mean that one can obtain a good fit of the distribution of $(1/n)(\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp}) - c_0)$ by the conditional distribution of $(1/n)(\mathbf{C}(\tilde{\mathbb{X}}_n) - \mathbb{E}[\mathbf{C}(\tilde{\mathbb{X}}_n)|\mathbb{X}_n])$ given $\mathbb{X}_n$. But as seen above, constructing a test based on the test statistic $(1/n)(\mathbf{C}(\mathbb{X}_n) - c_0)$ is impossible in a full distribution free context where the value of c_0 is

unknown.
Therefore one needs to find a quantity that is completely observable but whose mean is still null under $(\mathcal{H}_0)$. The statistic $\mathbf{U}$ introduced in Section 2.3.1 is suitable from this point of view. What one needs to check is whether the distribution of $\mathbf{U}(\mathbb{X}_n)$ under $(\mathcal{H}_0)$, that is of $\mathbf{U}(\mathbb{X}_n^{\perp\!\!\!\perp})$ (which has zero mean), is well approximated by the distribution of $\mathbf{U}\left(\tilde{\mathbb{X}}_n\right) - \mathbb{E}\left[\mathbf{U}\left(\tilde{\mathbb{X}}_n\right)|\mathbb{X}_n\right]$. For the trial-shuffling, since

$$\mathbf{U}\left(\mathbb{X}_n^{TS}\right) = \sum_{k=1}^{n} \varphi\left(X^1_{i^{TS}(k)}, X^2_{j^{TS}(k)}\right) - \frac{1}{n-1}\sum_{k\neq k'} \varphi\left(X^1_{i^{TS}(k)}, X^2_{j^{TS}(k')}\right),$$

one can easily see that because the couple $\left(i^{TS}(k), j^{TS}(k)\right)$ is drawn uniformly at random in the set of the (i,j)'s such that $i \neq j$ (set of cardinality $n(n-1)$),

$$\begin{aligned}
\mathbb{E}\left[\mathbf{U}\left(\mathbb{X}_n^{TS}\right)|\mathbb{X}_n\right] &= \frac{1}{n-1}\sum_{i\neq j}\varphi\left(X_i^1, X_j^2\right) - \frac{1}{n}\sum_{i,j}\varphi\left(X_i^1, X_j^2\right) \\
&= \frac{\hat{\mathbf{C}}_0\left(\mathbb{X}_n\right) - \mathbf{C}\left(\mathbb{X}_n\right)}{n} \\
&= -\frac{\mathbf{U}\left(\mathbb{X}_n\right)}{n}.
\end{aligned}$$

Hence the correct bootstrap statistic is

$$\tilde{\mathbf{U}}^{TS} = \tilde{\mathbf{U}}\left(\mathbb{X}_n^{TS}\right) = \mathbf{U}\left(\mathbb{X}_n^{TS}\right) + \frac{\mathbf{U}\left(\mathbb{X}_n\right)}{n}.$$

However similar computations show that the full bootstrap and the permutation satisfy

$$\mathbb{E}\left[\mathbf{U}\left(\mathbb{X}_n^{*}\right)|\mathbb{X}_n\right] = \mathbb{E}\left[\mathbf{U}\left(\mathbb{X}_n^{\Pi_n}\right)|\mathbb{X}_n\right] = 0,$$

so $\mathbf{U}\left(\mathbb{X}_n^{*}\right)$ and $\mathbf{U}\left(\mathbb{X}_n^{\Pi_n}\right)$ can be used directly.
Figure 2.6 shows the quality of approximation of the distribution of $\mathbf{U}\left(\mathbb{X}_n^{\perp\!\!\!\perp}\right)$ by the conditional distribution given the observation of either $\mathbf{U}^* = \mathbf{U}\left(\mathbb{X}_n^{*}\right)$ or $\mathbf{U}^\star = \mathbf{U}\left(\mathbb{X}_n^{\Pi_n}\right)$. The approximation is accurate under $(\mathcal{H}_0)$ but it is actually also accurate even if the observed sample is simulated under $(\mathcal{H}_1)$, which is in complete accordance with the mathematical consistence results in the Wasserstein metric proved in Chapter 1. The approximation is just as accurate for the recentered statistic $\tilde{\mathbf{U}}^{TS} = \mathbf{U}^{TS} + \mathbf{U}^{obs}/n$. Note that the difference between the conditional c.d.f. of $\tilde{\mathbf{U}}^{TS}$ and the one of $\mathbf{U}^{TS}$ is particularly visible under $(\mathcal{H}_1)$ when $X^1 = X^2$. Hence, as explained by the computations above, in a trial-shuffling approach, the recentered version leads to the correct bootstrap distribution. Note finally that this corroborates the previous intuition: the reason why the approximation works for $\mathbf{U}$ and not for $\mathbf{C}$ is exactly the same as for the naive approach of Figure 2.3. The centering is indeed random (here it can be viewed as $\mathbb{E}\left[\mathbf{C}(\tilde{\mathbb{X}}_n)\middle|\mathbb{X}_n\right]$) and one needs to take it into account to have a correct approximation.
Finally an extra simplification holds in the permutation case, which may seem very surprising. One can easily rewrite on the one hand,

$$\mathbf{U}\left(\mathbb{X}_n\right) = \left(1 + \frac{1}{n-1}\right)\mathbf{C}\left(\mathbb{X}_n\right) - \frac{1}{n-1}\sum_{i,j}\varphi\left(X_i^1, X_j^2\right)$$

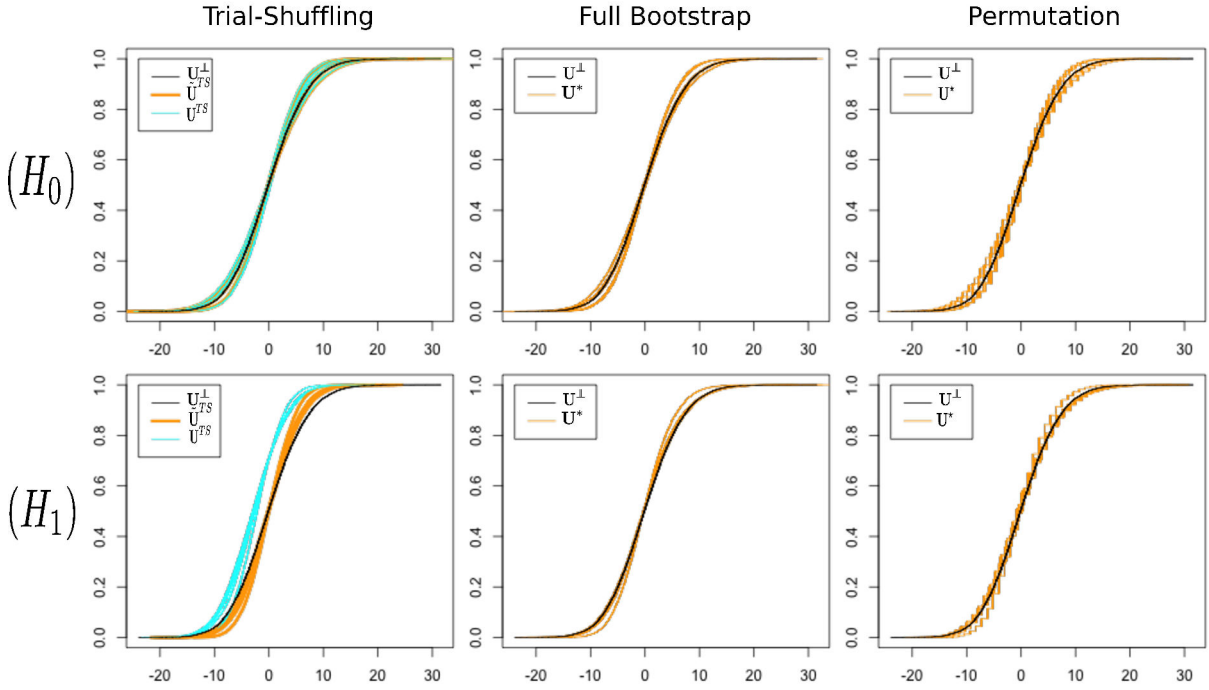


Figure 2.6 – Conditional distribution of $\mathbf{U}(\tilde{\mathbb{X}}_n)$ (or its recentered version $\tilde{\mathbf{U}}^{TS}$ for the trial-shuffling) given $\mathbb{X}_n$. C.d.f. of $\mathbf{U}^{\perp\!\!\!\perp} = \mathbf{U}\left(\mathbb{X}_n^{\perp\!\!\!\perp}\right)$ (in black), obtained by simulation as in Figure 2.4. For the first line, under $(\mathcal{H}_0)$, five observations of $\mathbb{X}_n^{\perp\!\!\!\perp}$ in the same set-up have been fixed and given these observations, the conditional c.d.f. of $\mathbf{U}^{TS} = \mathbf{U}\left(\mathbb{X}_n^{TS}\right)$, of $\tilde{\mathbf{U}}^{TS} = \mathbf{U}^{TS} + \mathbf{U}^{obs}/n$, of $\mathbf{U}^* = \mathbf{U}\left(\mathbb{X}_n^*\right)$ and of $\mathbf{U}^\star = \mathbf{U}\left(\mathbb{X}_n^{\Pi_n}\right)$ (in orange) have been obtained as in Figure 2.4. For the second line, five observations of $\mathbb{X}_n$, simulated under $(\mathcal{H}_1)$ with marginals equal to the ones of the first line but satisfying $X^1 = X^2$, have been simulated and conditional c.d.f. are obtained in the same way as above.

---

and, on the other hand, for the permutation sample

$$\mathbf{U}\left(\mathbb{X}_n^{\Pi_n}\right) = \left(1 + \frac{1}{n-1}\right)\mathbf{C}\left(\mathbb{X}_n^{\Pi_n}\right) - \frac{1}{n-1}\sum_{i,j}\varphi\left(X_i^1, X_j^2\right).$$

Note that the sum $\sum_{i,j}\varphi\left(X_i^1, X_j^2\right)$ is invariant by the action of the permutation. Hence if $u_t^\star$ denotes the quantile of order t of the conditional distribution of $\mathbf{U}\left(\mathbb{X}_n^{\Pi_n}\right)$ given $\mathbb{X}_n$ and if $c_t^\star$ denotes the quantile of order t of the conditional distribution of $\mathbf{C}\left(\mathbb{X}_n^{\Pi_n}\right)$ given $\mathbb{X}_n$, this very simple relationship holds

$$u_t^\star = \left(1 + \frac{1}{n-1}\right)c_t^\star - \frac{1}{n-1}\sum_{i,j}\varphi\left(X_i^1, X_j^2\right).$$

Hence the test that rejects $(\mathcal{H}_0)$ when $\mathbf{U}\left(\mathbb{X}_n\right) > u_{1-\alpha}^\star$ is exactly the one that rejects $(\mathcal{H}_0)$ when $\mathbf{C}\left(\mathbb{X}_n\right) > c_{1-\alpha}^\star$. Therefore despite the fact that the conditional distribution of $\mathbf{C}\left(\mathbb{X}_n^{\Pi_n}\right)$ is not close at all to the one of $\mathbf{C}\left(\mathbb{X}_n^{\perp\!\!\!\perp}\right)$, the test based on $\mathbf{C}$ works, because it is equivalent

to the test based on $\mathbf{U}$, for which the approximation of the conditional distribution works. Note however that this phenomenon happens only in the permutation approach, but not in the trial-shuffling or the full bootstrap approaches.

2.3.4 Practical testing procedures and p-values

From the considerations given above, five different tests may be investigated, the first one based on a purely asymptotic approach, and the four other ones based on bootstrap or permutation approaches, with critical values approximated through a Monte Carlo method. For each test, the corresponding p-values (that is the values of α for which the test passes from acceptance to rejection) are given.

The naive test (N) It consists in rejecting $(\mathcal{H}_0)$ when

$$\mathbf{Z}^{obs} \geq z_{1-\alpha}.$$

The corresponding p-value is given by

$$1 - \Phi\left(\mathbf{Z}^{obs}\right),$$

where Φ is the c.d.f. of a standard Gaussian distribution.

The Trial-Shuffling test, version C (TSC) It consists in rejecting $(\mathcal{H}_0)$ when

$$\mathbf{C}^{obs} \geq \hat{c}^{TS}_{1-\alpha},$$

where $\hat{c}^{TS}_{1-\alpha}$ is the empirical quantile of order $(1-\alpha)$ of the conditional distribution of $\mathbf{C}^{TS}$ given $\mathbb{X}_n$. This empirical quantile is estimated over B ($B = 10{,}000$ usually) realizations $\mathbf{C}^{TS}_1, ..., \mathbf{C}^{TS}_B$ given the observed sample $\mathbb{X}_n$. The corresponding p-value is given by

$$\frac{1}{B}\sum_{i=1}^{B} \mathbf{1}_{\mathbf{C}^{TS}_i \geq \mathbf{C}^{obs}}.$$

Despite the problems underlined in Section 2.3.3, we kept this test in the present study since it corresponds to the one programmed in [138] and since this bootstrap procedure is usually the one applied by neuroscientists.

The Trial-Shuffling test, version recentered U (TSU) It consists in rejecting $(\mathcal{H}_0)$ when

$$\mathbf{U}^{obs} \geq \hat{w}^{TS}_{1-\alpha},$$

where $\hat{w}^{TS}_{1-\alpha}$ is the empirical quantile of order $(1-\alpha)$ of the conditional distribution of $\tilde{\mathbf{U}}^{TS}$ (the correctly recentered statistic) given $\mathbb{X}_n$. This empirical quantile and the corresponding p-value are obtained in a similar way as for the above (TSC), based on B realizations $\tilde{\mathbf{U}}^{TS}_1, ..., \tilde{\mathbf{U}}^{TS}_B$ of $\tilde{\mathbf{U}}\left(\mathbb{X}^{TS}_n\right)$ given $\mathbb{X}_n$.

The Full Bootstrap test, version U (FBU) It consists in rejecting $(\mathcal{H}_0)$ when

$$\mathbf{U}^{obs} \geq \hat{u}^{*}_{1-\alpha},$$

where $\hat{u}^{*}_{1-\alpha}$ is the empirical quantile of order $(1-\alpha)$ of the conditional distribution of $\mathbf{U}^*$ given $\mathbb{X}_n$. This empirical quantile and the corresponding p-value are obtained in a similar way as for the above (TSC), based on B realizations $\mathbf{U}^*_1, ..., \mathbf{U}^*_B$ of $\mathbf{U}\left(\mathbb{X}^*_n\right)$ given $\mathbb{X}_n$.

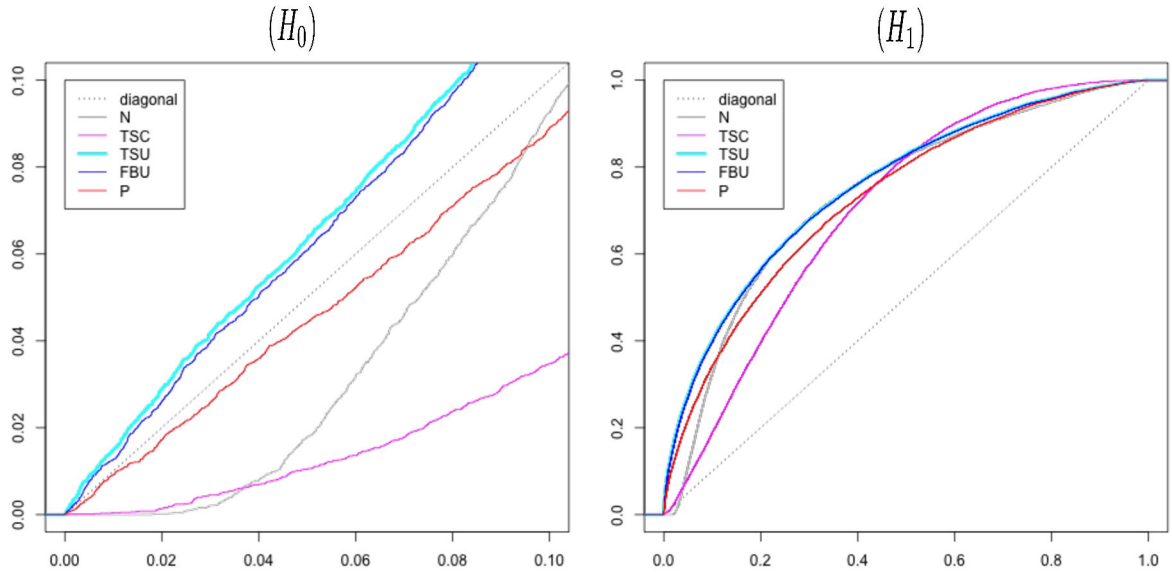


Figure 2.7 – Distribution of the p-values for the different tests. C.d.f. under both $(\mathcal{H}_0)$ and $(\mathcal{H}_1)$ of the p-values for the five tests: naive (N), Trial-Shuffling version **C** (TSC), Trial-Shuffling version **U** (TSU), Full Bootstrap version **U** (FBU), and Permutation (P). Under $(\mathcal{H}_0)$, the c.d.f. are obtained by simulations done as in Figure 2.4; the c.d.f. are then plotted only for small p-values (≤ 0.1). Under $(\mathcal{H}_1)$, the couple (X^1, X^2) is constructed by injection [71, 170], that is as $(N^1 \cup N^{inj}, N^2 \cup N^{inj})$ where (N^1, N^2) are two independent Poisson processes of firing rate 27 Hz on a window of length 0.1s and where N^{inj} is a common Poisson process of firing rate 3Hz independent of (N^1, N^2); once again, 20 i.i.d. trials are simulated 10,000 times to obtain the corresponding c.d.f. with $\delta = 0.01$s.

The permutation test (P) The reader may think that it should consist in rejecting $(\mathcal{H}_0)$ when

$$\mathbf{C}^{obs} \geq \hat{c}^{\star}_{1-\alpha},$$

where $\hat{c}^{\star}_{1-\alpha}$ is the empirical quantile of order $(1-\alpha)$ of the conditional distribution of $\mathbf{C}^{\star}$ given $\mathbb{X}_n$. But the test by permutation is in fact directly defined by its p-value, which is slightly different here, equal to

$$\frac{1}{B+1}\left(1 + \sum_{i=1}^{B} \mathbf{1}_{\mathbf{C}^{\star}_i \geq \mathbf{C}^{obs}}\right),$$

where $\mathbf{C}^{\star}_1, ..., \mathbf{C}^{\star}_B$ are B realizations of $\mathbf{C}\left(\mathbb{X}_n^{\Pi_n}\right)$ given $\mathbb{X}_n$. The permutation test then consists in rejecting $(\mathcal{H}_0)$ when this p-value is less than α. Indeed, such a permutation test, with such a slightly different version of p-value, has been proved to be exactly of level α, whatever B (see [156]).

Note however that such a slight correction does not work for full bootstrap or trial-shuffling approaches, where the tests are only guaranteed to be asymptotically of level α.

Saying that a test rejects at level α is equivalent to saying that its p-value is less than α. If a test is exactly of level α for any α in $(0, 1)$, the c.d.f. of its p-values should therefore be smaller that the one of a uniform variable (that is the diagonal). Between several tests

with this guarantee, the less conservative one is the one for which the c.d.f. of its p-values is the closest to the diagonal. The left-hand side of Figure 2.7 shows the c.d.f. under $(\mathcal{H}_0)$ of the corresponding p-values for the five considered testing procedures and focuses on small p-values, which are the only ones usually involved in testing, to highlight the main differences between the five methods. For the chosen value of n ($n = 20$), the c.d.f. of the (TSU) and (FBU) p-values are almost identical and above the diagonal, meaning that the corresponding tests do not guarantee the level. On the contrary, the c.d.f. of the (N) and (TSC) p-values are clearly under the diagonal and far from it, meaning that the corresponding tests are too conservative. As guaranteed by [156], the permutation approach guarantees the level of the test: the c.d.f. of the (P) p-values is also under the diagonal, but much closer to the diagonal than the one of the (N) and (TSC) p-values.
Furthermore, the behavior of the c.d.f. of the p-values under $(\mathcal{H}_1)$ gives an indication of the power of the test: the highest the c.d.f. of the p-values, the most powerful the corresponding test. Hence among the tests that guarantee the level, the permutation test (P) is the most powerful one. Note that other simulations in more various cases have been performed in Chapter 1 leading to the same conclusion.
In the sequel, the focus is therefore on the permutation approach, keeping also the trial-shuffling version $\mathbf{C}$ approach as a variant of the method developed in [138].

2.4 Multiple tests

2.4.1 Description of the complete multiple testing algorithm

To detect precise locations of dependence periods that can be matched to some experimental or behavioral events, it is classical to consider a family of windows $\mathcal{W}$ of cardinal K, which is a collection of potentially overlapping intervals $[a, b]$ covering the whole interval $[0, T]$ on which trials have been recorded [72, 170]. Then, some independence tests are implemented on each window of the collection. Here we propose a complete algorithm which takes into account the multiplicity of the tests, and which moreover enables to see if the coincidence count is significantly too large or too small on each window as in [170].

Permutation UE algorithm

Fix a real number q in $(0, 0.5)$ and an integer B larger than 2.

- Do in parallel for all window $W = [a, b]$ in $\mathcal{W}$:
  * Extract the points of the X_i^1's and X_i^2's in $[a, b]$.
  * For all (i, j) in $\{1, ..., n\}^2$, compute $a_{i,j} = \varphi_\delta^{coinc}\left(X_i^1, X_j^2\right)$ over $[a, b]$ thanks to the **delayed coincidence count algorithm**.
  * Draw at random B i.i.d. permutations $\Pi_n^{\mathbf{b}}$, $1 \leq \mathbf{b} \leq B$, and compute $\mathbf{C}^{\mathbf{b}} = \sum_i a_{i,\Pi_n^{\mathbf{b}}(i)}$.
  * Compute also $\mathbf{C}^{obs} = \sum_i a_{i,i}$.
  * Return $p_W^+ = \frac{1}{B+1}\left(1 + \sum_{\mathbf{b}=1}^B \mathbf{1}_{\mathbf{C}^{\mathbf{b}} \geq \mathbf{C}^{obs}}\right)$ and $p_W^- = \frac{1}{B+1}\left(1 + \sum_{\mathbf{b}=1}^B \mathbf{1}_{\mathbf{C}^{\mathbf{b}} \leq \mathbf{C}^{obs}}\right)$.
- Perform the procedure of [16] on the set of the above $2K$ p-values:
  * Sort the p-values $p^{(1)} \leq ... \leq p^{(2K)}$.
  * Find $k = \max\{l \ ; \ p^{(l)} \leq lq/(2K)\}$.
  * Return the (W, ϵ_W)'s for which W is associated with one of the p-values $p^{(l)}$ for $l \leq k$, with $\epsilon_W = 1$ if $p_W^+ \leq p^{(k)}$, so the coincidence count is significantly too large on W, and $\epsilon_W = -1$ if $p_W^- \leq p^{(k)}$, so the coincidence count is significantly too small on W.

This algorithm corresponds to a slight variation of the multiple testing step of [170], but adapted to not necessarily symmetric distributions [5]. In several applications, neuroscientists are interested in detecting dependence periods for which the coincidence count is only significantly too large. In this case, one can use the restricted set of the p^+_W's. Then if the considered windows are disjoint and if the spike trains are Poisson processes that are not necessarily stationary, the False Discovery Rate (FDR) [6] of the above multiple testing procedure is mathematically proved[7] to be controlled by q for any $B \geq 2$.
The code has been parallelized in `C++` and interfaced with `R`.
The full corresponding `R`-package is still a work in progress but actual codes are available at `https://github.com/ybouret/neuro-stat`.

2.4.2 Comparison on simulations

Two sets of simulations have been performed. The first one, namely Experiment 1, combines different point processes encountered in the literature (homogeneous, and inhomogeneous Poisson processes, Hawkes processes), and different kinds of dependences. It is described in Figure 2.8.**A**. The second one, namely Experiment 2, is consisted of simple independent homogeneous Poisson processes on the whole interval $[0, 2]$, as described in Table 2.1. The corresponding results are described in Table 2.1 and one run of simulation of the Permutation UE method is presented in Figure 2.8. Four methods have been compared:

- the MTGAUE method of [170] which assumes both processes to be homogeneous Poisson processes,
- the Trial-Shuffling, version **C** (TSC) which corresponds to the method of [138], which has been programmed with the delayed coincidence count described above and which has not been corrected for multiplicity.
- the same as above but corrected by Benjamini and Hochberg procedure (TSC + BH),
- the Permutation UE approach described above.

The permutation approach always guarantees an FDR less than the prescribed level of 0.05 whereas MTGAUE does not when the homogeneous Poisson assumption fails (as in Experiment 1, see Figure 2.8). The classical trial-shuffling method (where dependence detection occurs each time the p-value is less than 0.05) seems to have comparable results in terms of both FDR and False Non Discovery Rate (FNDR) on Experiment 1 but fails to control the FDR on the most basic situation, namely purely independent processes (Experiment 2). Adding a Benjamini-Hochberg step of selection of p-values to the trial-shuffling makes it more robust but at the price of a much larger FNDR with respect to the Permutation UE method, fact which is consistent with the conservativeness shown in Figure 2.7.

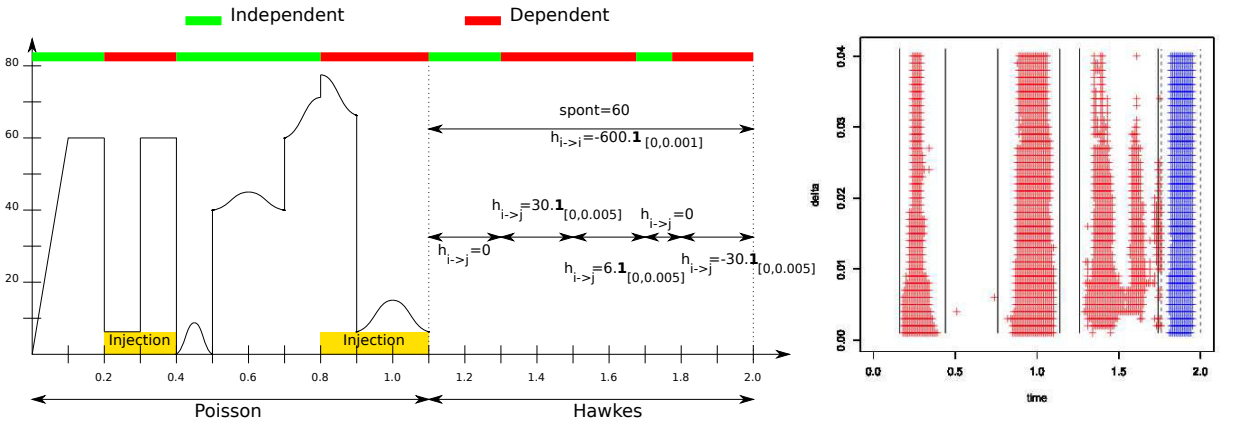
2.4.3 Comparison on real data

Behavioral procedure The data used in this chapter to test the dependence detection ability of the four methods were already partially published in previous experimental studies

[5] Note in particular that for a fixed W, one cannot have both $p^+_W < 0.5$ and $p^-_W < 0.5$ and therefore, if a W is detected, it can only be because of one of the two situations, $p^+_W \leq p^{(k)}$ or $p^-_W \leq p^{(k)}$, which cannot happen simultaneously.

[6] see [170] or Table 2.1 for a precise definition

[7] The p^+_W's are independent random variables such that $\mathbb{P}_{\perp\!\!\!\perp}\left(p^+_W \leq \alpha\right) \leq \alpha$ for all α in $[0, 1]$ [17, 156].



A: Description of Experiment 1 **B: Result of the permutation method**

Figure 2.8 – Multiple tests. 2.8.**A**: description of Experiment 1. In the Poisson part, the intensity of both Poisson processes is plotted. The injection component corresponds to the part of a shared Poisson process which is injected in both processes corresponding to X^1 and X^2, as explained in Figure 2.7. In the Hawkes part (see [170] for a complete description, or Section 0.6.4, Definition 0.6.4 in the introduction of this thesis), formulas for the spontaneous parameters and both self interaction $h_{i\to i}$ and cross interaction $h_{i\to j}$ functions are given. 2.8.**B**: results of the Permutation UE method ($B = 10{,}000$, $q = 0.05$) performed on 191 overlapping windows of the form $[a, a + 0.1]$ for a in $\{0, 0.01, ..., 1.9\}$ on one run of simulation for 50 trials of Experiment 1. A red (resp. blue) cross is represented at the center of the window when it is detected by a p^+_W (resp. p^-_W). Each horizontal line corresponds to a different δ in $\{0.001, 0.002, ...0.04\}$. The black vertical lines delimit the regions where the independence hypothesis is not satisfied: plain for positive dependence (that is where $\mathbf{C}^{obs}$ should be too large), and dashed for negative dependence (that is where $\mathbf{C}^{obs}$ should be too small). The dotted vertical line separates the region of high (on the left) and low (on the right) dependence in the Hawkes positive dependence part.

[150, 64, 151] and also used in [170]. These data were collected on a 5-year-old male Rhesus monkey who was trained to perform a delayed multidirectional pointing task. The animal sat in a primate chair in front of a vertical panel on which seven touch-sensitive light-emitting diodes were mounted, one in the center and six placed equidistantly (60 degrees apart) on a circle around it. The monkey had to initiate a trial by touching and then holding with the left hand the central target. After a fix delay of 500ms, the preparatory signal (PS) was presented by illuminating one of the six peripheral targets in green. After a delay of either 600ms or 1200ms, selected at random with various probability, it turned red, serving as the response signal and pointing target. During the first part of the delay, the probability p_{resp} for the response signal to occur at $(500 + 600)$ms $= 1.1$s was 0.3. Once this moment passed without signal occurrence, the conditional probability for the signal to occur at $(500 + 600 + 600)$ms $= 1.7$s changed to 1. The monkey was rewarded by a drop of juice after each correct trial. Reaction time (RT) was defined as the release of the central target. Movement time (MT) was defined as the touching of the correct peripheral target.

Recording technique Signals recorded from up to seven microelectrodes (quartz insulated platinum-tungsten electrodes, impedance: $2 - 5$MΩ at 1,000Hz) were amplified and band-

	Accepted.	Rejected.	Total
Independ.	U	V	m_0
Depend.	T	S	$m - m_0$
Total	$m - R$	R	m

	Experiment 1		Experiment 2	
	FDR	FNDR	FDR	FNDR
MTGAUE	0.10	0.17	0.04	0
TSC	0.01	0.26	0.25	0
TSC + BH	0	0.32	0	0
P	0.01	0.23	0.02	0

Table 2.1 – False Discovery and Non Discovery Rates. On the left-hand side, the classical table for multiple testing adapted to our dependence framework, with a total number of tests $m = 2K$. On the right-hand side, estimated FDR and FNDR over 1,000 runs, FDR being defined by $\mathbb{E}\left[(V/R)\mathbf{1}_{R>0}\right]$ and FNDR being defined by $\mathbb{E}\left[(T/(m-R))\mathbf{1}_{m-R>0}\right]$. Experiment 1 is described in Figure 2.8, Experiment 2 consists in two independent homogeneous Poisson processes of firing rate 60 Hz on $[0, 2]$. The set of windows is as in Figure 2.8. There are 50 trials and $\delta = 0.01$s. MTGAUE is the method described in [170] with $q = 0.05$. (TSC) is the trial-shuffling method with Monte Carlo approximation ($B = 10{,}000$) and the selected windows are the ones whose p-value are less than 0.05. (TSC+BH) is the same method, except that the multiplicity of the tests is corrected by a Benjamini-Hochberg procedure ($q = 0.05$). (P) corresponds to the Permutation UE method ($B = 10{,}000$, $q = 0.05$).

pass filtered from 300Hz to 10kHz. Using a window discriminator, spikes from only one single neuron per electrode were then isolated. Neuronal data along with behavioral events (occurrences of signals and performance of the animal) were stored on a PC for off-line analysis with a time resolution of 10kHz.

In the following study, only trials where the response signal (RS) occurs at 1.7s are considered. The expected signal (ES) corresponds to an eventually expected but not confirmed signal, that is at 1.2s. Only the pair 13 of the previous data set is considered here, as it was one of the main two examples already treated in [170].

The results are presented in Figure 2.9. The (TSC+BH) method does not detect anything and is therefore not presented. The Permutation UE method detects less windows than both (MTGAUE) and (TSC) methods, but the detected windows are still in adequation with the experimental or behavioral events. The above simulation study let us think that the extra detections of both (MTGAUE) and (TSC) may be false positives, since both methods do not control the FDR as well as the Permutation UE method.

2.5 Conclusion

After describing a fast algorithm to compute the delayed coincidence count, showing that this notion can be used in practice for any surrogate data method in place of the binned coincidence count, we have focused on distribution free methods to test the independence between two simultaneously recorded spike trains. Though they are here presented with the delayed coincidence count, all these distribution free methods could be applied to any coincidence count if desired.

Once the coincidence count $\mathbf{C}$ chosen, we have first introduced an empirical quantity or statistic $\mathbf{U}$ whose distribution is centered under the independence hypothesis ($\mathcal{H}_0$). In the spirit of [170] but in a distribution free manner, a first naive method consists in performing a Gaussian approximation of the distribution of $\mathbf{U}$ under ($\mathcal{H}_0$). This method suffers from a not very

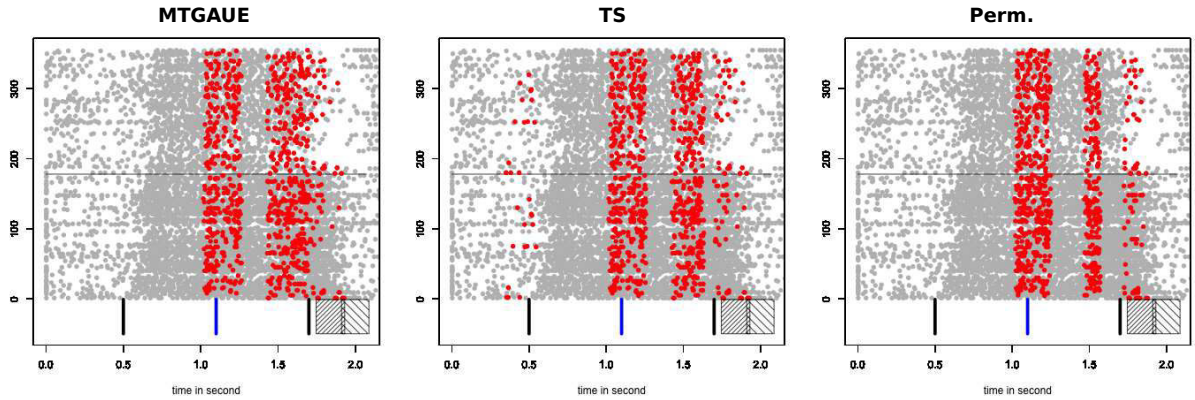


Figure 2.9 – Raster plots of the pair of neurons 13. In red the Unitary Events where the coincidence count is significantly too large for the three methods (MTGAUE, TSC and P) presented in Table 2.1, with $\delta = 0.02$s and $B = 10{,}000$. No interval was detected for a significantly too small coincidence count. Signs on bottom corresponds to behavioral events. The first black vertical bar corresponds to the preparatory signal (PS), the blue vertical bar to the expected signal (ES), the second black vertical bar to the response signal (RS).

sharp approximation when the number of trials n is small (see Figure 2.2). Moreover the approximation is clearly not valid when the observed sample does not satisfy $(\mathcal{H}_0)$ (see the last line of Figure 2.3).

We then turned to bootstrap and permutation methods. One of the most well-known bootstrap method in the neuroscience literature is the trial-shuffling [138, 137]. It is usually based on a resampling approach directly applied on the coincidence count itself, namely $\mathbf{C}$. The other two investigated methods (full bootstrap and permutation) are well-known in the statistics literature for independence testing between real-valued random vectors [84, 155] and have been applied to point processes that are modeling here simultaneously recorded spike trains in Chapter 1. These last two methods are usually applied on centered statistics.

One of the main message of the present work is that applying resampling methods to non centered statistics, as $\mathbf{C}$, does not lead to a correct approximation of the distribution of $\mathbf{C}$ under $(\mathcal{H}_0)$ (see the second line of Figure 2.4). This phenomenon, combined with observations in the more classical framework of the naive test on Figure 2.3, leads us to think about a centering defect. Once the methods are applied on correct centered statistics ($\mathbf{U}$ or $\tilde{\mathbf{U}}$ depending on the chosen resampling method), they all three outperform the naive approach. The approximation is better under $(\mathcal{H}_0)$ for small value of n (first line of Figure 2.3 and first line of Figure 2.6) and is still accurate when the observed sample does not satisfy $(\mathcal{H}_0)$ (last line of Figure 2.3 and second line of Figure 2.6).

From an algorithmic point of view, all the corresponding p-values are evaluated thanks to a Monte Carlo algorithm and a program which interfaces `R` and `C++`, thus making the running time fast and the use easy. Pipa and Grün [137] have given an exact algorithm when the trial-shuffling is applied on the coincidence count $\mathbf{C}$ directly. It is a very elegant algorithm using the fact that $\mathbf{C}$ is an integer that can take a small number of values. Unfortunately the same gain is not really possible for $\mathbf{U}$ which is not an integer and which can take much more values. Moreover this exact algorithm is quite long with respect to the Monte Carlo algorithm when the number of simulations is 10,000 (as used in the present work) and one can see on

the bottom left of Figure 2.4 that the difference between both results (Monte Carlo and exact algorithms) is not detectable at first glance.
A more precise study of the Monte Carlo approximated bootstrap p-values shows that for a small number of trials, the trial-shuffling and the full bootstrap methods, even applied to a correctly centered statistic, do not provide tests of prescribed level. On the contrary, the permutation method, thanks to an adequate version of its p-values [156], allows for a precise control of the level. The classical trial-shuffling method based on the non centered quantity $\mathbf{C}$ and the naive approach also both lead to a precise control of the level but in a more conservative way (see the right-hand side of Figure 2.7). This is also showed by the behavior of the p-values under the alternative, p-values that are smaller for the permutation approach than for the other two methods (see the right-hand side of Figure 2.7).
Finally, we decided to combine the delayed coincidence count, which is much more precise than the binned coincidence count [170, 72] with the permutation approach, and to apply the obtained independence testing procedure to several windows of detection simultaneously. The final proposed method consists in combining the individual tests with the approach of [16] to correct for the multiplicity of the tests. Parallel programming is used to treat each window in an independent manner. This new algorithm named Permutation UE is completely distribution free. It better controls the False Discovery Rate than MTGAUE [170] or the classical trial-shuffling method applied on $\mathbf{C}$ (see Table 2.1, methods (MTGAUE) and (TSC)). Moreover, it does not suffer from conservativeness as the trial-shuffling method applied on $\mathbf{C}$, once the multiplicity of the tests is taken into account (see Table 2.1, method (TSC+BH)). On real data, the results are similar to existing methods (MTGAUE, TSC) except for some detections that disappear but that are likely to be false positive thanks to the present simulation study.
To conclude, we introduce in this chapter the Permutation UE method, which is a Unitary Events method based on delayed coincidence count and on an evaluation of p-values via a distribution free Monte Carlo approximated permutation approach. This method does not suffer from any loss in synchrony detection as the binned coincidence count [67], is distribution free and in this sense upgrades [170]. Moreover the algorithm is fast and parallelized, and despite using a Monte Carlo scheme, it can guarantee the single tests to be of the prescribed level and the multiple test to control the FDR in a non-asymptotic manner, therefore outperforming the trial-shuffling method [138, 137] in terms of both mathematical caution and computing time, when compared with the exact algorithm described in [137]. Finally it is still sufficiently sensitive to detect reasonable features on real data sets. The only drawback is that it can only work for pairs of neurons. The definition of delayed coincidence count for more than two neurons has been recently introduced in [35], but the combination of this notion with a bootstrap or a permutation approach is still an open question.

Chapter 3

Concentration inequalities for randomly permuted sums

Initially motivated by the study of the non-asymptotic properties of non-parametric tests based on permutation methods as introduced in Chapters 1 and 4, some concentration inequalities for uniformly permuted sums are derived from the fundamental inequalities for random permutations of Talagrand. The idea is to first obtain a rough inequality for the square root of the permuted sum, and then, iterate the previous analysis and plug this first inequality to obtain a general concentration of permuted sums around their median. Then, concentration inequalities around the mean are deduced, thanks to Ledoux's trick. This method allows us to obtain a Bernstein-type inequality. In particular, one recovers the Gaussian behavior of such permuted sums under classical conditions encountered in the literature.

The work presented in this chapter has been done without any collaboration. However, thanks to a completely different approach based on martingale inequalities, comparable results (with better constants) have been independently obtained at the same time by Bernard Delyon during the Fall 2014. A joint work with him to present those inequalities in a unified manner and with some applications is still in progress.

Contents

3.1 Introduction and motivation 166
3.2 Bernstein-type concentration inequalities for permuted sums . . 172
3.2.1 Concentration of permuted sums of non-negative numbers 172
3.2.2 Concentration of permuted sums in the general case 174
3.3 Proofs 175
3.3.1 Proof of Lemma 3.2.1 175
3.3.2 Proof of Proposition 3.2.1 177
3.3.3 Proof of Proposition 3.2.2 180
3.3.4 Proof of Corollary 3.2.1 182
3.3.5 Proof of Theorem 3.2.1 182
3.3.6 Proof of Corollary 3.2.2 183
3.4 Appendix: a non-asymptotic control of the second kind error . . 184
3.4.1 A first condition ensuing from Chebychev's inequality 185
3.4.2 Control of the variance in the general case 186
3.4.3 Control of the critical value based on Hoeffding's approach 188

3.4.4 Control of the quantile of the critical value 191
3.4.5 A new condition ensuing from Hoeffding's approach 192
3.4.6 An alternative proof of the consistency of the permutation test . . . 193

3.1 Introduction and motivation

This chapter presents concentration inequalities for randomly permuted sums defined by $Z_n = \sum_{i=1}^n a_{i,\Pi_n(i)}$, where $\{a_{i,j}\}_{1\leq i,j\leq n}$ are real numbers, and Π_n is a uniformly distributed random permutation of the set $\{1,\dots,n\}$. Initially motivated by hypothesis testing in the non-parametric framework (see [178] for instance), such sums have been largely studied from an asymptotic point of view in the literature. A first combinatorial central limit theorem is proved by Wald and Wolfowitz in [178], in the particular case when the real numbers $a_{i,j}$ are of a product form $b_i \times c_j$, under strong assumptions that have been released for instance by Noether [128]. Then, Hoeffding obtains stronger results in such product case, and generalizes those results to not necessarily product type real terms $a_{i,j}$ in [83]. For instance, he proves (see [83, Theorem 3]) that, if

$$\lim_{n\to+\infty} \frac{\frac{1}{n}\sum_{1\leq i,j\leq n} d_{i,j}^r}{\left(\frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2\right)^{r/2}} = 0, \quad \text{for some } r>2, \tag{3.1.1}$$

where

$$d_{i,j} = a_{i,j} - \frac{1}{n}\sum_{k=1}^n a_{k,j} - \frac{1}{n}\sum_{l=1}^n a_{i,l} + \frac{1}{n^2}\sum_{k,l=1}^n a_{k,l}, \tag{3.1.2}$$

then the distribution of $Z_n = \sum_{i=1}^n a_{i,\Pi_n(i)}$ is asymptotically normal, that is, for all x in $\mathbb{R}$,

$$\lim_{n\to+\infty} \mathbb{P}\left(Z_n - \mathbb{E}\left[Z_n\right] \leq x\sqrt{\mathrm{Var}(Z_n)}\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-\frac{y^2}{2}}dy.$$

He also proposes a stronger (in the sense that it implies (3.1.1)), but simpler condition in [83, Theorem 3], precisely

$$\frac{\max_{1\leq i,j\leq n}\{|d_{i,j}|\}}{\sqrt{\frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2}} \underset{n\to+\infty}{\longrightarrow} 0, \tag{3.1.3}$$

under which such an asymptotic Gaussian limit holds. Similar results have been obtained later, for instance by Motoo [127], under the following Lindeberg-type condition that is for all $\varepsilon > 0$,

$$\lim_{n\to+\infty} \sum_{1\leq i,j\leq n} \left(\frac{d_{i,j}}{d}\right)^2 \mathbb{1}_{\left|\frac{d_{i,j}}{d}\right|>\varepsilon} = 0, \tag{3.1.4}$$

where $d^2 = n^{-1}\sum_{1\leq i,j\leq n} d_{i,j}^2$. In particular, he proves in [127] that such Lindeberg-type condition is weaker than Hoeffding's ones in the sense that (3.1.4) is implied by (3.1.1) (and thus by (3.1.3)). A few years later, Hájek [74] proves in the product case, that the condition (3.1.4) is in fact necessary. A simpler proof of the sufficiency of the Lindeberg-type condition is given by Schneller [163] based on Stein's method.

Afterwards, the next step was to study the convergence of the conditional distribution when the terms $a_{i,j}$ in the general case, or $b_i \times c_j$ in the product case, are random. Notably, Dwass

studies in [49] the limit of the randomly permuted sum in the product case, where only the c_j's are random, and proves that the conditional distribution given the c_j's converges almost surely (a.s.) to a Gaussian distribution. Then, Shapiro and Hubert [165] generalized this study to weighted U-statistics of the form $\sum_{i\neq j} b_{i,j}h(X_i, X_j)$ where the X_i's are independent and identically distributed (i.i.d.) random variables. In a first time, they show some a.s. asymptotic normality of this statistic. In a second time, they complete Jogdeo's [101] work in the deterministic case, proving asymptotic normality of permuted statistics based on the previous weighted U-statistic. More precisely, they consider the rank statistic $\sum_{i\neq j} b_{i,j}h(X_{R_i}, X_{R_j})$, where R_i is the rank of V_i in a sample $V_1, \dots, V_n$ of i.i.d. random variables with a continuous distribution function. In particular, notice that considering such rank statistics is equivalent to considering uniformly permuted statistics. In Chapter 1 of this thesis, in collaboration with Y. Bouret, M. Fromont and P. Reynaud-Bouret, we generalize previous combinatorial central limit theorems to permuted sums of non-i.i.d. random variables $\sum_{i=1}^n Y_{i,\Pi_n(i)}$, for particular forms of random variables $Y_{i,j}$, and particular dependences recalled below. The main difference with the previous results comes from the fact that the random variables $Y_{i,j}$ are not necessarily exchangeable.
Hence, the asymptotic behavior of permuted sums have been vastly investigated in the literature, allowing to deduce good properties for permutation tests based on such statistics, like the asymptotic size, or the power (see for instance [155] or Chapter 1). Yet, such results are purely asymptotic, while, in many application fields, such as neurosciences for instance as described in Chapters 1 or 2, few exploitable data are available. Hence, such asymptotic results may not be sufficient. This is why a non-asymptotic approach is preferred here.

Before investigating the non-asymptotic literature, let us recall the main motivation of this chapter. Consider the independence testing framework introduced in Chapter 1. Let $\mathcal{X}$ represent either the set of all countable subsets of $[0,1]$ or possibly a more general separable set. Given an i.i.d. n-sample $\mathbb{X}_n = (X_1, \dots, X_n)$, where each X_i is a couple (X_i^1, X_i^2) in $\mathcal{X}^2$ with distribution P of marginals P^1 and P^2, we aim at testing

$$\text{the null hypothesis } (\mathcal{H}_0) \text{ "} P = P^1 \otimes P^2 \text{" against the alternative } (\mathcal{H}_1) \text{ "} P \neq P^1 \otimes P^2 \text{".}$$

The considered test statistic corresponding to the *Linear case* in Chapter 1, is defined by

$$T_\delta(\mathbb{X}_n) = \frac{1}{n-1}\left(\sum_{i=1}^n \varphi_\delta(X_i^1, X_i^2) - \frac{1}{n}\sum_{i,j=1}^n \varphi_\delta(X_i^1, X_j^2)\right), \tag{3.1.5}$$

where φ_δ is a measurable real-valued function on $\mathcal{X}^2$ potentially depending on some parameter δ. Notice that here, as non-asymptotic results are expected, the scaling term $\sqrt{n}$ in Chapter 1 is dropped.
The critical value of the test is obtained from the permutation approach, inspired by Hoeffding [84], and Romano [155]. Let Π_n be a uniformly distributed random permutation of $\{1, \dots, n\}$ independent of $\mathbb{X}_n$ and consider the permuted test statistic defined by

$$T_\delta(\mathbb{X}_n^{\Pi_n}) = \frac{1}{n-1}\left(\sum_{i=1}^n \varphi_\delta(X_i^1, X_{\Pi_n(i)}^2) - \frac{1}{n}\sum_{i,j=1}^n \varphi_\delta(X_i^1, X_j^2)\right),$$

which is the test statistic computed on the permuted sample $\mathbb{X}_n^{\Pi_n} = (X_1^{\Pi_n}, \dots X_n^{\Pi_n})$ obtained from permuting only the second coordinates, where precisely $X_i^{\Pi_n} = (X_i^1, X_{\Pi_n(i)}^2)$ for all

$1 \leq i \leq n$. Then, the critical value of the upper-tailed test, denoted by $q_{1-\alpha}(\mathbb{X}_n)$, is the $(1-\alpha)$-quantile of the conditional distribution of the permuted statistic $T_\delta(\mathbb{X}_n^{\Pi_n})$ given the sample $\mathbb{X}_n$. More precisely, given $\mathbb{X}_n$, if

$$T_\delta^{(1)}(\mathbb{X}_n) \leq T_\delta^{(2)}(\mathbb{X}_n) \leq \cdots \leq T_\delta^{(n!)}(\mathbb{X}_n)$$

denote the ordered values of all the permuted test statistic $T_\delta(\mathbb{X}_n^{\pi_n})$, when π_n describes the set of all permutations of $\{1, \ldots, n\}$, then the critical value is equal to

$$q_{1-\alpha}(\mathbb{X}_n) = T_\delta^{(n!-\lfloor n!\alpha \rfloor)}(\mathbb{X}_n). \tag{3.1.6}$$

The corresponding test rejects the null hypothesis when $T_\delta(\mathbb{X}_n) > q_{1-\alpha}(\mathbb{X}_n)$, here denoted by

$$\Delta_\alpha(\mathbb{X}_n) = \mathbb{1}_{T_\delta(\mathbb{X}_n) > q_{1-\alpha}(\mathbb{X}_n)}. \tag{3.1.7}$$

Notice that, as in Chapter 1, one may consider the lower-tailed test considering the conditional α-quantile given $\mathbb{X}_n$, that is $q_\alpha^-(\mathbb{X}_n) = T_\delta^{(\lfloor n!\alpha \rfloor)}(\mathbb{X}_n)$, and reject the null hypothesis when $T_\delta(\mathbb{X}_n) < q_\alpha^-(\mathbb{X}_n)$ but only the upper-tailed test is considered here as a motivation.
In Chapter 1, the asymptotic properties of such test are studied. Based on the Gaussian limit distribution of the permuted test statistic stated in Theorem 1.4.1 under either the null hypothesis or any reasonable alternatives, and other mild conditions, the test is proved to be asymptotically of prescribed size, and power equal to one (see Theorem 1.4.2). Yet, as explained above, such purely asymptotic properties may be insufficient when applying these tests in neuroscience for instance. Moreover, the delicate choice of the parameter δ is a real question, especially, in neuroscience, where it has some biological meaning, as mentioned in Chapters 1 and 2. A possible approach to overcome this issue is to aggregate several tests for different parameters δ, and reject independence if at least one of them does. In particular, this approach should give us information on how to choose this parameter. Yet, to do so, non-asymptotic controls are necessary.
From a non-asymptotic point of view, the test is proved to be non-asymptotically of prescribed level in the introduction of this thesis. Hence, remains the non-asymptotic control of the second kind error rate, that is the probability of wrongly accepting the null hypothesis. In the spirit of [57, 58, 160], the idea is to study the uniform separation rates, as done in the special case of real-valued random variables in Chapter 4. From now on, consider an alternative P satisfying $(\mathcal{H}_1)$, and an i.i.d. sample $\mathbb{X}_n$ from such distribution P. As only upper-tailed tests are considered here, assume moreover that the alternative satisfies $\mathbb{E}\left[\varphi_\delta(X_1^1, X_1^2)\right] > \mathbb{E}\left[\varphi_\delta(X_1^1, X_2^2)\right]$, that is $\mathbb{E}\left[T_\delta(\mathbb{X}_n)\right] > 0$. The initial step is to find some condition on P guaranteeing the control of the second kind error rate, namely $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0)$, by a prescribed value $\beta > 0$. Intuitively, since the expectation of the test statistic $\mathbb{E}\left[T_\delta(\mathbb{X}_n)\right]$ is equal to zero under the null hypothesis, the test should be more efficient in rejecting $(\mathcal{H}_0)$ for large values of this expectation. So, the aim is to find conditions of the form $\mathbb{E}\left[T_\delta(\mathbb{X}_n)\right] \geq s$ for some threshold s to be determined. Yet, one of the main difficulties here comes from the randomness of the critical value. The idea, as in [57], is thus to introduce $q_{1-\beta/2}^\alpha$ the $(1-\beta/2)$-quantile of the critical value $q_{1-\alpha}(\mathbb{X}_n)$ and deduce from Chebychev's inequality (see the Appendix section 3.4.1 of this chapter), that the second kind error rate is controlled by β as soon as

$$\mathbb{E}\left[T_\delta(\mathbb{X}_n)\right] \geq q_{1-\beta/2}^\alpha + \sqrt{\frac{2}{\beta}\operatorname{Var}(T_\delta(\mathbb{X}_n))}. \tag{3.1.8}$$

Usually, the goal in general minimax approaches is to express, for well-chosen functions φ_δ, some distance between the alternative P and the null hypothesis $(\mathcal{H}_0)$ thanks to $\mathbb{E}\left[T_\delta(\mathbb{X}_n)\right]$

for which minimax lower-bounds are known (see for instance [57, 58]). The objective is then to control, up to a constant, such distance (and in particular each term in the right-hand side of (3.1.8)) by the minimax rate of independence testing with respect to such distance on well-chosen classes of alternatives, in order to prove the optimality of the method from a theoretical point of view. The interested reader could refer to Chapter 4 for more details about this kind of development in the density case. It is not in the scope of the present chapter to develop such minimax theory in the general case, but to provide some general tools providing some sharp control of each term in the right-hand side of (3.1.8) which consists in a very first step of this approach. It is shown, in the Appendix section 3.4.2 of this chapter, that the variance term can be upper bounded, up to a multiplicative constant, by $n^{-1}(\mathbb{E}\left[\varphi_\delta^2(X_1^1, X_1^2)\right] + \mathbb{E}\left[\varphi_\delta^2(X_1^1, X_2^2)\right])$. Hence, the challenging part relies in the quantile term. At this point, we explored several ideas.

The first one is based on the non-asymptotic control of the critical value obtained in the Appendix section 3.4.3 of this chapter (see equation (3.4.3)), following Hoeffding's idea (see [84, Theorem 2.1]), that leads to the condition

$$\mathbb{E}\left[T_\delta(\mathbb{X}_n)\right] \geq \frac{4}{\sqrt{\alpha}}\sqrt{\frac{2}{\beta}\frac{\mathbb{E}\left[\varphi_\delta(X_1^1, X_1^2)^2\right] + \mathbb{E}\left[\varphi_\delta(X_1^1, X_2^2)^2\right]}{n}}. \tag{3.1.9}$$

The proof of this result is detailed in the Appendix section 3.4.5 of this chapter. Notice that this approach provides an alternative proof of the consistency of this permutation test, as also detailed in the Appendix section 3.4.6 of this chapter. Yet, this result may not be sharp enough, especially in α. Indeed, as explained above, the next step consists in aggregating several tests for different values of the parameter δ in a purpose of adaptivity. Generally, when aggregating tests, as in multiple testing methods, the multiplicity of the tests has to be taken into account. In particular, the single prescribed level of each individual test should be corrected. Several corrections exist, such as the Bonferroni one, which consists in dividing the global desired level α by the number of tests M. Yet, for such correction, the lower-bound in (3.1.9) comes with a cost in $\sqrt{M}$, which is too large to provide optimal rates. Even with more sophisticate corrections than the Bonferroni one, such as the one considered in Chapter 4, the control by a term of order $\alpha^{-1/2}$ is too large, since classically in the literature (see, e.g., [57, 58, 160]), the dependence in α should be of the order of $\sqrt{\ln(1/\alpha)}$. Hence, the bound ensuing from this first track being not sharp enough, the next idea was to investigate other non-asymptotic approaches for permuted sums.

Such approaches have also been studied in the literature. For instance, Ho and Chen [80] obtain non-asymptotic Berry-Esseen type bounds in the $\mathbb{L}^p$-distance between the cumulative distribution function (c.d.f.) of the standardized permuted sum of i.i.d. random variables and the c.d.f. of the normal distribution, based on Stein's method. In particular, they obtain the rate of convergence to a normal distribution in $\mathbb{L}^p$-distance under Lindeberg-type conditions. Then, Bolthausen [25] proposes a different approach, also based on Stein's method allowing to extend Ho and Chen's results in the non-identically distributed case. More precisely, he obtains bounds in the $\mathbb{L}^\infty$-distance in the non-random case. In particular, in the deterministic case (which can easily be generalized to random cases), considering the notation introduced above, he obtains the following non-asymptotic bound:

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(Z_n - \mathbb{E}\left[Z_n\right] \leq x\sqrt{\mathrm{Var}(Z_n)}\right) - \Phi_{0,1}(x)\right| \leq \frac{C}{n\sqrt{\mathrm{Var}(Z_n)}^3}\sum_{i,j=1}^{n}|d_{i,j}|^3,$$

where C is an absolute constant, and $\Phi_{0,1}$ denotes the standard normal distribution function. In particular, when applying this result to answer our motivation by considering random variables $\varphi_\delta(X_i^1, X_j^2)$ instead of the deterministic terms $a_{i,j}$, and working conditionally on the sample $\mathbb{X}_n$, the permuted statistic $T_\delta(\mathbb{X}_n^{\Pi_n})$ corresponds to $(n-1)^{-1}(Z_n - \mathbb{E}[Z_n])$. Therefore, the previous inequality implies that, for all t in $\mathbb{R}$,

$$\mathbb{P}\left(T_\delta\left(\mathbb{X}_n^{\Pi_n}\right) \leq t \middle| \mathbb{X}_n\right) \leq \left[1 - \Phi_{0,1}\left(\frac{t}{\sqrt{\mathrm{Var}\left(T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\middle|\mathbb{X}_n\right)}}\right)\right] + \frac{C}{n(n-1)^{2/3}\sqrt{\mathrm{Var}\left(T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\middle|\mathbb{X}_n\right)}^3}\sum_{i,j}|D_{i,j}|^3, \tag{3.1.10}$$

where $D_{i,j}$ denotes

$$\varphi_\delta(X_i^1, X_j^2) - \frac{1}{n}\sum_{l=1}^n \varphi_\delta(X_i^1, X_l^2) - \frac{1}{n}\sum_{k=1}^n \varphi_\delta(X_k^1, X_j^2) + \frac{1}{n^2}\sum_{k,l=1}^n \varphi_\delta(X_k^1, X_l^2).$$

Yet, by definition of conditional quantiles, the critical value $q_{1-\alpha}(\mathbb{X}_n)$ is the smallest value of t such that $\mathbb{P}(T_\delta(\mathbb{X}_n) \leq t|\mathbb{X}_n) \leq \alpha$. Hence, considering (3.1.10), one can easily make the first term of the sum in the right-hand side of the inequality as small as one wants by choosing t large enough. However, the second term being fixed, nothing guarantees that the upper-bound in (3.1.10) can be constrained to be smaller than α. Thus, this result cannot be applied in order to control non-asymptotically the critical value.

Concentration inequalities seem thus to be adequate here, as they provide sharp non-asymptotic results, with usually exponentially small controls which leads to the desired logarithmic dependency in α, as mentioned above. They have been vastly investigated in the literature, and the interested reader can refer to the books of Ledoux [114], Massart [123], or the very recent one of Boucheron, Lugosi, and Massart [26] for some overall reviews. Generally, they provide precise tail bounds for well-behaved functions or sums of independent random variables, such as, for instance, the classical Bernstein inequality recalled in Theorem 0.5.1 in the introduction of this thesis. Yet, permuted sums are sums of dependent variables.
The work in this chapter is based on the pioneering work of Talagrand (see [168] for a review) who investigates the concentration of measure phenomenon for product measures. Of main interest here, he proved the following inequality for random permutations in [168, Theorem 5.1].

Theorem 3.1.1 (Talagrand, 1995)**.** *Denote by $\mathfrak{S}_n$ the set of all permutations of $\{1, \ldots, n\}$. Define for any subset $A \subset \mathfrak{S}_n$, and permutation $\pi_n \in \mathfrak{S}_n$,*

$$U_A(\pi_n) = \left\{s \in \{0,1\}^n \ ; \ \exists \tau \in A \text{ such that } \forall 1 \leq i \leq n, \ s_i = 0 \implies \tau(i) = \pi_n(i)\right\}.$$

Then, consider $V_A(\pi_n) = \mathrm{ConvexHull}\left(U_A(\pi_n)\right)$, and

$$f(A, \pi_n) = \min\left\{\sum_{i=1}^n v_i^2 \ ; \ v = (v_i)_{1\leq i\leq n} \in V_A(\pi_n)\right\}.$$

Then, if P_n denotes the uniform distribution on $\mathfrak{S}_n$,

$$\int_{\mathfrak{S}_n} e^{\frac{1}{16}f(A,\pi_n)} dP_n(\pi_n) \leq \frac{1}{P_n(A)}.$$

Therefore, by Markov's inequality, for all $t > 0$,

$$P_n\left(\pi_n \ ; \ f(A,\pi_n) \geq t^2\right) \leq \frac{e^{-t^2/16}}{P_n(A)}. \tag{3.1.11}$$

This result on random permutations is fundamental, and is a key point to many other non-asymptotic works on random permutations. Among them emerges McDiarmid's article [124] in which he derives from Talagrand's inequality, exponential concentration inequalities around the median for randomly permuted functions of the observation under Lipschitz-type conditions and applied to randomized methods for graph coloring. More recently, Adamczak et al. obtained in [2] some concentration inequality under convex-Lipschitz conditions when studying the empirical spectral distribution of random matrices. In particular, they proves the following Theorem (precisely [2, Theorem 3.1]).

Theorem 3.1.2 (Adamczak, Chafai and Wolff, 2014)**.** *Consider* $x_1, \ldots, x_n$ *in* $[0,1]$ *and let* $\varphi : [0,1]^n \to \mathbb{R}$ *be an L-Lipschitz convex function. Let* Π_n *be a random uniform permutation of the set* $\{1,\ldots,n\}$ *and denote* $Z = \varphi\big(x_{\Pi_n(1)}, \ldots, x_{\Pi_n(n)}\big)$*. Then, there exists some positive absolute constant* c *such that, for all* $t > 0$,

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq t) \leq 2\exp\left(-\frac{ct^2}{L^2}\right).$$

Yet, the Lispchitz assumptions may be very restrictive. It is not clear that the particular functions considered in the application fields satisfy such assumption, and it is definitively not the case when considering for instance Haar wavelets as done in the case of real-valued random variables in Chapter 4, since they are not even continuous. Hence, the idea is to exploit the attractive form of a sum of our test statistic. Based on Stein's method, initially introduced to study the Gaussian behavior of sums of dependent random variables, Chatterjee studies permuted sums of non-negative numbers in [33]. He obtains in [33, Proposition 1.1] the following first Bernstein-type concentration inequality for non-negative terms around the mean.

Theorem 3.1.3 (Chatterjee, 2007)**.** *Let* $\{a_{i,j}\}_{1\leq i,j\leq n}$ *be a collection of numbers from* $[0,1]$*. Let* $Z_n = \sum_{i=1}^n a_{i,\Pi_n(i)}$*, where* Π_n *is drawn from the uniform distribution over the set of all permutations of* $\{1,\ldots,n\}$*. Then, for any* $t \geq 0$,

$$\mathbb{P}(|Z_n - \mathbb{E}[Z_n]| \geq t) \leq 2\exp\left(-\frac{t^2}{4\mathbb{E}[Z_n] + 2t}\right). \tag{3.1.12}$$

Notice that because of the expectation term in the right-hand side of (3.1.12), the link with Hoeffding's combinatorial central limit theorem (for instance) is not so clear. In this chapter, this result is sharpened in the sense that this expectation term is replaced by a variance term, allowing us to provide a non-asymptotic version of such combinatorial central limit theorem. This result is moreover generalized to any real numbers (not necessarily non-negative).

The present work is organized as follows. In Section 3.2 are formulated the main results. Section 3.2.1 is devoted to the permuted sums of non-negative numbers. Based on Talagrand's result, a first rough concentration inequality for the square root of permuted sum is obtained in Lemma 3.2.1. Then by iterating the previous analysis and plugging this first inequality, a general concentration of permuted sums around their median is obtained in Proposition

3.2.1. Finally, the concentration inequality of Proposition 3.2.2 around the mean is deduced thanks to Ledoux' trick in [114, Proposition 1.8]. In Section 3.2.2, the previous inequalities are generalized to general permuted sums of not necessarily non-negative terms. The proofs are detailed in Section 3.3. Finally, the Appendix section 3.4 contains the non-asymptotic control of the second kind error rate of the permutation test introduced in this introduction as a motivation.

3.2 Bernstein-type concentration inequalities for permuted sums

Let us first introduce some general notation. In the sequel, denote by $\mathfrak{S}_n$ the set of permutations of $\{1, 2, \dots, n\}$. For all collection of real numbers $\{a_{i,j}\}_{1\le i,j\le n}$, and for each π_n in $\mathfrak{S}_n$, consider the permuted sum

$$Z_n(\pi_n) = \sum_{i=1}^{n} a_{i,\pi_n(i)}.$$

Let Π_n be a random uniform permutation in $\mathfrak{S}_n$, and $Z_n := Z_n(\Pi_n)$. Denote $\mathrm{med}\,(Z_n)$ its median, that is which satisfies

$$\mathbb{P}(Z_n \ge \mathrm{med}\,(Z_n)) \ge 1/2 \quad \text{and} \quad \mathbb{P}(Z_n \le \mathrm{med}\,(Z_n)) \ge 1/2.$$

This study is divided in two steps. The first one is restrained to non-negative terms. The second one extends the previous results to general terms, based on a trick involving both non-negative and negative parts.

3.2.1 Concentration of permuted sums of non-negative numbers

In the present section, the collection of numbers $\{a_{i,j}\}_{1\le i,j\le n}$ is assumed to be non-negative. The proof of the concentration inequality around the median in Proposition 3.2.1 needs a preliminary step which is presented in Lemma 3.2.1. It provides concentration inequality for the square root of the sum. It allows us then by iterating the same argument, and plugging the obtained inequality to the square root of the sum of the squares, namely $\sqrt{\sum_{i=1}^n a_{i,\Pi_n(i)}^2}$, to be able to sharpen Chatterjee's concentration inequality (3.1.12).

Lemma 3.2.1. *Let $\{a_{i,j}\}_{1\le i,j\le n}$ be a collection of non-negative numbers, and Π_n be a random uniform permutation in $\mathfrak{S}_n$. Consider $Z_n = \sum_{i=1}^n a_{i,\Pi_n(i)}$. Then, for all $t > 0$,*

$$\mathbb{P}\left(\sqrt{Z_n} \ge \sqrt{\mathrm{med}\,(Z_n)} + t\sqrt{\max_{1\le i,j\le n}\{a_{i,j}\}}\right) \le 2e^{-t^2/16}, \tag{3.2.1}$$

and

$$\mathbb{P}\left(\sqrt{Z_n} \le \sqrt{\mathrm{med}\,(Z_n)} - t\sqrt{\max_{1\le i,j\le n}\{a_{i,j}\}}\right) \le 2e^{-t^2/16}. \tag{3.2.2}$$

In particular, one obtains the following two-sided concentration for the square root of a randomly permuted sum of non-negative numbers,

$$\mathbb{P}\left(\left|\sqrt{Z_n} - \sqrt{\mathrm{med}\,(Z_n)}\right| > t\sqrt{\max_{1\le i,j\le n}\{a_{i,j}\}}\right) \le 4e^{-t^2/16}.$$

The idea of the proof is the same that the one of Adamczak et al. in [2, Theorem 3.1], but with a sum instead of a convex Lipschitz function. In a similar way, it is based on Talagrand's inequality for random permutations recalled in Theorem 3.1.1 in the introduction of this Chapter.

In the following are presented two concentration inequalities in the non-negative case; the first one around the median, and the second one around the mean. It is well known that both are equivalent up to constants, but here, both are detailed in order to give the order of magnitude of the constants. The transition from the median to the mean is based on Ledoux' trick in the proof of [114, Proposition 1.8] allowing to reduce exponential concentration inequalities around any constant m (corresponding in our case to $\mathrm{med}\,(Z_n)$) to similar inequalities around the mean. This trick consists in using the exponentially fast decrease around m to upper bound the difference between m and the mean.

Proposition 3.2.1. *Let $\{a_{i,j}\}_{1\leq i,j\leq n}$ be a collection of non-negative numbers and Π_n be a random uniform permutation in $\mathfrak{S}_n$. Consider $Z_n = \sum_{i=1}^n a_{i,\Pi_n(i)}$. Then, for all $x>0$,*

$$\mathbb{P}\left(|Z_n - \mathrm{med}\,(Z_n)| > \sqrt{\mathrm{med}\left(\sum_{i=1}^n a_{i,\Pi_n(i)}^2\right)x} + x\max_{1\leq i,j\leq n}\{a_{i,j}\}\right) \leq 8\exp\left(\frac{-x}{16}\right). \quad (3.2.3)$$

Since in many applications, the concentration around the mean is more adapted, the following proposition shows that one may obtain a similar behavior around the mean, at the cost of drastic constants.

Proposition 3.2.2. *Let $\{a_{i,j}\}_{1\leq i,j\leq n}$ be a collection of non-negative numbers, and Π_n be a random uniform permutation in $\mathfrak{S}_n$. Consider $Z_n = \sum_{i=1}^n a_{i,\Pi_n(i)}$.*
Then, for all $x>0$,

$$\mathbb{P}\left(|Z_n - \mathbb{E}\,[Z_n]| \geq 2\sqrt{\left(\frac{1}{n}\sum_{i,j=1}^n a_{i,j}^2\right)x} + \max_{1\leq i,j\leq n}\{a_{i,j}\}\,x\right) \leq 8e^{16\pi}\exp\left(-\frac{x}{16}\right). \quad (3.2.4)$$

This concentration inequality is called a Bernstein-type inequality restricted to non-negative sums, due to its resemblance to the standard Bernstein inequality, as recalled in Theorem 0.5.1 in the introduction of this thesis. The main difference here lies in the fact that the random variables in the sum are not independent. Moreover, this inequality implies a more popular form of Bernstein's inequality stated in Corollary 3.2.1.

Corollary 3.2.1. *With the same notation and assumptions as in Proposition 3.2.2, for all $t>0$,*

$$\mathbb{P}(|Z_n - \mathbb{E}\,[Z_n]| \geq t) \leq 8e^{16\pi}\exp\left(\frac{-t^2}{16\left(4\frac{1}{n}\sum_{i,j=1}^n a_{i,j}^2 + 2\max_{1\leq i,j\leq n}\{a_{i,j}\}\,t\right)}\right). \quad (3.2.5)$$

Comment: Recall Chatterjee's result in [33, Proposition 2.1], quoted in Theorem 3.1.3, which can easily be rewritten with our notation, and for a collection of non-negative numbers not necessarily in $[0,1]$, by

$$\forall t>0, \quad \mathbb{P}(|Z_n - \mathbb{E}\,[Z_n]| \geq t) \leq 2\exp\left(\frac{-t^2}{4M_a\frac{1}{n}\sum_{i,j=1}^n a_{i,j} + 2M_a t}\right),$$

where M_a denotes the maximum $\max_{1\leq i,j\leq n}\{a_{i,j}\}$. Up to the constants, the inequality in (3.2.4) is sharper thanks to the quadratic term since $\sum_{i,j=1}^n a_{i,j}^2 \leq M_a \sum_{i,j=1}^n a_{i,j}$ always holds.

3.2.2 Concentration of permuted sums in the general case

In this section, the collection of numbers $\{a_{i,j}\}_{1\leq i,j\leq n}$ is no longer assumed to be non-negative. The following general concentration inequality for randomly permuted sums directly derives from Proposition 3.2.2.

Theorem 3.2.1. *Let $\{a_{i,j}\}_{1\leq i,j\leq n}$ be a collection of any real numbers, and Π_n be a random uniform permutation in $\mathfrak{S}_n$. Consider $Z_n = \sum_{i=1}^n a_{i,\Pi_n(i)}$. Then, for all $x>0$,*

$$\mathbb{P}\left(|Z_n - \mathbb{E}[Z_n]| \geq 2\sqrt{2\left(\frac{1}{n}\sum_{i,j=1}^n a_{i,j}^2\right)x} + 2\max_{1\leq i,j\leq n}\{|a_{i,j}|\}\,x\right) \leq 16e^{16\pi}\exp\left(-\frac{x}{16}\right). \tag{3.2.6}$$

Once again, the obtained inequality is a Bernstein-type inequality. Moreover, it is also possible to obtain a more popular form of Bernstein-type inequalities applying the same trick based on the non-negative and the negative parts from Corollary 3.2.1.

Corollary 3.2.2. *With the same notation as in Theorem 3.2.1, for all $t>0$,*

$$\mathbb{P}(|Z_n - \mathbb{E}[Z_n]| \geq t) \leq 16e^{16\pi}\exp\left(\frac{-t^2}{256\left(\mathrm{Var}(Z_n) + \max_{1\leq i,j\leq n}\{|a_{i,j}|\}\,t\right)}\right).$$

Comments: One recovers a Gaussian behavior of the centered permuted sum obtained by Hoeffding in [83, Theorem 3] under the same assumptions. Indeed, in the proof of Corollary 3.2.2, one obtains the following intermediate result (see (3.3.14)), that is

$$\mathbb{P}(|Z_n - \mathbb{E}[Z_n]| \geq t) \leq 16e^{16\pi}\exp\left(\frac{-t^2}{64\left(4\frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2 + \max_{1\leq i,j\leq n}\{|d_{i,j}|\}\,t\right)}\right),$$

where the $d_{i,j}$'s are defined in (3.1.2). Yet, $\mathrm{Var}(Z_n) = \frac{1}{n-1}\sum_{i,j=1}^n d_{i,j}^2$ (see [83, Theorem 2]). Hence, applying this inequality to $t = x\sqrt{\mathrm{Var}(Z_n)} \geq x\sqrt{n^{-1}\sum_{i,j=1}^n d_{i,j}^2}$ for $x>0$ leads to

$$\mathbb{P}\left(|Z_n - \mathbb{E}[Z_n]| \geq x\sqrt{\mathrm{Var}(Z_n)}\right) \leq 16e^{16\pi}\exp\left(\frac{-x^2}{256\left(1+\frac{\max_{1\leq i,j\leq n}\{|d_{i,j}|\}}{\sqrt{\frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2}}x\right)}\right),$$

Hence, under Hoeffding's simpler condition (3.1.3), namely

$$\lim_{n\to+\infty}\frac{\max_{1\leq i,j\leq n} d_{i,j}^2}{\frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2} = 0,$$

one recovers,

$$\lim_{n\to+\infty}\mathbb{P}\left(|Z_n - \mathbb{E}[Z_n]| \geq x\sqrt{\mathrm{Var}(Z_n)}\right) \leq 16e^{16\pi}e^{-x^2/256},$$

which is a Gaussian tail that is, up to constants, close in spirit to the one obtained by Hoeffding in [83, Theorem 3].

3.3 Proofs

3.3.1 Proof of Lemma 3.2.1

Sketch of proof. From now on, fix $t > 0$. Recall the notation introduced by Talagrand in Theorem 3.1.1. As mentioned in the introduction of this thesis (see Section 0.5.3), the main purpose of these notation is to introduce some notion of distance between a permutation π_n in $\mathfrak{S}_n$ and a subset A of $\mathfrak{S}_n$. To do so, the idea is to reduce the set of interest to a simpler one, that is $[0,1]^n$, by considering

$$U_A(\pi_n) = \{s \in \{0,1\}^n \ ; \ \exists \tau \in A \text{ such that } \forall 1 \leq i \leq n, \ s_i = 0 \implies \tau(i) = \pi_n(i)\}.$$

One may notice that the permutation π_n belongs to A if and only if 0 belongs to the set $U_A(\pi_n)$. Hence, the corresponding distance between the permutation π_n and the set A is coded by the distance between 0 and the set $U_A(\pi_n)$ and thus defined by

$$f(A, \pi_n) = \min \left\{ \sum_{i=1}^{n} v_i^2 \ ; \ v = (v_i)_{1 \leq i \leq n} \in V_A(\pi_n) \right\},$$

where $V_A(\pi_n) = \text{ConvexHull}\,(U_A(\pi_n))$. One may notice in particular that A contains π_n if and only if the distance $f(A, \pi_n) = 0$.
The global frame of the proof of Lemma 3.2.1 (and also Proposition 3.2.1) relies on the following steps. The first step consists in proving that

$$\mathbb{P}\left(\sqrt{Z} \geq \sqrt{C_A} + t \sqrt{\max_{1 \leq i,j \leq n} \{a_{i,j}\}} \right) \leq \frac{e^{-t^2/16}}{\mathbb{P}(Z \in A)}, \tag{3.3.1}$$

for some subset A of $\mathfrak{S}_n$ of the shape $A = \{\tau \in \mathfrak{S}_n ; Z(\tau) \leq C_A\}$ for some constant C_A to be chosen later. For this purpose, since Talagrand's inequality for random permutations (see Theorem 3.1.1) provides that

$$\mathbb{P}\big(f(A, \Pi_n) \geq t^2\big) \leq \frac{e^{-t^2/16}}{\mathbb{P}(\Pi_n \in A)},$$

it is sufficient to prove that

$$\mathbb{P}\big(f(A, \Pi_n) \geq t^2\big) \geq \mathbb{P}\left(\sqrt{Z} \geq \sqrt{C_A} + t \sqrt{\max_{1 \leq i,j \leq n} \{a_{i,j}\}} \right),$$

to obtain (3.3.1). To do so, the idea, as in [2], is to show that the assertion $f(A, \Pi_n) < t^2$ implies that $\sqrt{Z} < \sqrt{C_A} + t\sqrt{\max_{1 \leq i,j \leq n} \{a_{i,j}\}}$, and to conclude by contraposition.
Then, the two following steps consist in choosing appropriate constants C_A in (3.3.1) depending on the median of Z, such that both $\mathbb{P}\left(\sqrt{Z} \geq \sqrt{C_A} + t\sqrt{\max_{1 \leq i,j \leq n} \{a_{i,j}\}} \right)$ and $\mathbb{P}(Z \in A)$ are greater than $1/2$, in order to control both probabilities

$$\mathbb{P}\left(\sqrt{Z} \geq \sqrt{\text{med}\,(Z)} + t \sqrt{\max_{1 \leq i,j \leq n} \{a_{i,j}\}} \right) \text{ and } \mathbb{P}\left(\sqrt{Z} \leq \sqrt{\text{med}\,(Z)} - t \sqrt{\max_{1 \leq i,j \leq n} \{a_{i,j}\}} \right)$$

respectively in (3.2.1) and (3.2.2).

First step: preliminary study. Assume $f(A,\Pi_n) < t^2$. Then, by definition of the distance f, there exists some $s^1,\dots,s^m$ in $U_A(\Pi_n)$, and some non-negative weights $p_1,\dots,p_m$ satisfying $\sum_{j=1}^m p_j = 1$ such that

$$\sum_{i=1}^n \left[\left(\sum_{j=1}^m p_j s_i^j \right)^2 \right] < t^2.$$

For each $1 \le j \le m$, since s^j belongs to $U_A(\Pi_n)$, one may consider a permutation τ_j in A associated to s^j (that is satisfying $s_i^j = 0 \implies \tau_j(i) = \Pi_n(i)$). Then, since the $a_{i,j}$ are non-negative, and from the Cauchy-Schwartz inequality,

$$\begin{aligned}
Z - \sum_{j=1}^m p_j Z(\tau_j) &= \sum_{i=1}^n \sum_{j=1}^m p_j \left(a_{i,\Pi_n(i)} - a_{i,\tau_j(i)} \right) \\
&= \sum_{i=1}^n \sum_{j=1}^m p_j \left(a_{i,\Pi_n(i)} - a_{i,\tau_j(i)} \right) s_i^j \\
&\le \sum_{i=1}^n \left[\left(\sum_{j=1}^m p_j s_i^j \right) a_{i,\Pi_n(i)} \right] \\
&\le \sqrt{\sum_{i=1}^n \left(\sum_{j=1}^m p_j s_i^j \right)^2} \sqrt{\sum_{i=1}^n a_{i,\Pi_n(i)}^2} \\
&< t \sqrt{\max_{1\le i,j\le n} \{a_{i,j}\}} \sqrt{Z}.
\end{aligned}$$

Thus, as the τ_j are in $A = \{\tau; Z(\tau) \le C_A\}$,

$$Z < C_A + t \sqrt{\max_{1\le i,j\le n} \{a_{i,j}\}} \sqrt{Z}.$$

Therefore, by solving the second-order polynomial in $\sqrt{Z}$ above, one obtains

$$\sqrt{Z} < \frac{t\sqrt{\max_{1\le i,j\le n}\{a_{i,j}\}} + \sqrt{t^2 \max_{1\le i,j\le n}\{a_{i,j}\} + 4C_A}}{2} \le t \sqrt{\max_{1\le i,j\le n} \{a_{i,j}\}} + \sqrt{C_A}.$$

Finally, by contraposition,

$$\mathbb{P}\left(\sqrt{Z} \ge \sqrt{C_A} + t \sqrt{\max_{1\le i,j\le n} \{a_{i,j}\}} \right) \le \mathbb{P}\left(f(A,\Pi_n) \ge t^2 \right),$$

which, combined with (3.1.11) of Theorem 3.1.1 provides (3.3.1).

Second step: proof of (3.2.1). Taking $C_A = \operatorname{med}(Z)$ guarantees $\mathbb{P}(Z \in A) \ge 1/2$ and thus, (3.3.1) provides (3.2.1).

Third step: proof of (3.2.2). Taking $C_A = \left(\sqrt{\text{med}\,(Z)} - t\sqrt{\max_{1\le i,j\le n}\{a_{i,j}\}}\right)^2$ implies

$$\mathbb{P}\left(\sqrt{Z} \ge \sqrt{C_A} + t\sqrt{\max_{1\le i,j\le n}\{a_{i,j}\}}\right) = \mathbb{P}\left(\sqrt{Z} \ge \sqrt{\text{med}\,(Z)}\right) = \mathbb{P}(Z \ge \text{med}\,(Z)) \ge \frac{1}{2}.$$

So finally, again by (3.3.1),

$$\begin{aligned}\mathbb{P}\left(\sqrt{Z} \le \sqrt{\text{med}\,(Z)} - t\sqrt{\max_{1\le i,j\le n}\{a_{i,j}\}}\right) &= \mathbb{P}(Z \in A)\\ &\le \frac{e^{-t^2/16}}{\mathbb{P}\left(\sqrt{Z} \ge \sqrt{C_A} + t\sqrt{\max_{1\le i,j\le n}\{a_{i,j}\}}\right)}\\ &\le 2e^{-t^2/16},\end{aligned}$$

which ends the proof of the Lemma.

3.3.2 Proof of Proposition 3.2.1

From now on, fix $x > 0$, and consider $t = x^2$. This proof is again based on Talagrand's inequality for random permutations, combined with (3.2.1) in Lemma 3.2.1. It follows exactly the same progression as in the proof of Lemma 3.2.1; the preliminary step consists in working with subsets $A \subset \mathfrak{S}_n$ of the form $A = \{\tau \in \mathfrak{S}_n \ ; \ Z(\tau) \le C_A\}$ for some constant C_A, in order to obtain for all $v > 0$,

$$\mathbb{P}\left(Z \ge C_A + t\left(\sqrt{\text{med}\left(\sum_{i=1}^n a^2_{i,\Pi_n(i)}\right)} + v\max_{1\le i,j\le n}\{a_{i,j}\}\right)\right) \le \frac{e^{-t^2/16}}{\mathbb{P}(Z \in A)} + 2e^{-v^2/16}. \quad (3.3.2)$$

The second and third step consist in picking up a well-chosen constant C_A and a well-chosen $v > 0$ in order to obtain respectively

$$\mathbb{P}\left(Z \ge \text{med}\,(Z) + t\left(\sqrt{\text{med}\left(\sum_{i=1}^n a^2_{i,\Pi_n(i)}\right)} + (t \vee C_0)\max_{1\le i,j\le n}\{a_{i,j}\}\right)\right) \le 4e^{-t^2/16}, \quad (3.3.3)$$

and

$$\mathbb{P}\left(Z \le \text{med}\,(Z) - t\left(\sqrt{\text{med}\left(\sum_{i=1}^n a^2_{i,\Pi_n(i)}\right)} + (t \vee C_0)\max_{1\le i,j\le n}\{a_{i,j}\}\right)\right) \le 4e^{-t^2/16}, \quad (3.3.4)$$

where $C_0 = 4\sqrt{\ln(8)}$. The final step combines (3.3.3) and (3.3.4) in order to prove (3.2.3).

First step: preliminary study. Let $A = \{\tau \in \mathfrak{S}_n \ ; \ Z(\tau) \le C_A\}$ with C_A a general constant, and fix $v > 0$. Assume, this time, that both

$$f(A, \Pi_n) < t^2 \quad \text{and} \quad \sqrt{\sum_{i=1}^n a^2_{i,\Pi_n(i)}} < \sqrt{\text{med}\left(\sum_{i=1}^n a^2_{i,\Pi_n(i)}\right)} + v\max_{1\le i,j\le n}\{a_{i,j}\}. \quad (3.3.5)$$

Then, as in the preliminary study of the proof of Lemma 3.2.1, from the first assumption in (3.3.5), there exists some $s^1, \dots, s^m$ in $U_A(\Pi_n)$, and some non-negative weights $p_1, \dots, p_m$ satisfying $\sum_{j=1}^m p_j = 1$ such that

$$\sum_{i=1}^{n} \left[\left(\sum_{j=1}^{m} p_j s_i^j \right)^2 \right] < t^2.$$

For each $1 \leq j \leq m$, consider τ_j in A associated to s^j, that is a permutation τ_j in A satisfying $s_i^j = 0 \implies \tau_j(i) = \Pi_n(i)$. Then, combining the Cauchy-Shwartz inequality with the second assumption in (3.3.5) leads to

$$\begin{aligned} Z - \sum_{j=1}^{m} p_j Z(\tau_j) &= \sum_{i=1}^{n} \sum_{j=1}^{m} p_j \left(a_{i,\Pi_n(i)} - a_{i,\tau_j(i)} \right) s_i^j \\ &\leq \sum_{i=1}^{n} \left[\left(\sum_{j=1}^{m} p_j s_i^j \right) a_{i,\Pi_n(i)} \right] \\ &\leq \sqrt{\sum_{i=1}^{n} \left(\sum_{j=1}^{m} p_j s_i^j \right)^2} \sqrt{\sum_{i=1}^{n} a_{i,\Pi_n(i)}^2} \\ &< t \left(\sqrt{\operatorname{med} \left(\sum_{i=1}^{n} a_{i,\Pi_n(i)}^2 \right)} + v \max_{1 \leq i,j \leq n} \{a_{i,j}\} \right). \end{aligned}$$

Notice that here, the reasoning begins exactly as in the proof of Lemma 3.2.1. Yet, the second assumption in (3.3.5), which can be controlled thanks to that lemma, allows us to sharpen the inequality. Thus, as the τ_j are in $A = \{\tau; Z(\tau) \leq C_A\}$,

$$Z < C_A + t \left(\sqrt{\operatorname{med} \left(\sum_{i=1}^{n} a_{i,\Pi_n(i)}^2 \right)} + v \max_{1 \leq i,j \leq n} \{a_{i,j}\} \right). \tag{3.3.6}$$

Hence, by contraposition of (3.3.5) $\implies$ (3.3.6), one obtains

$$\begin{aligned} &\mathbb{P}\left(Z \geq C_A + t \left(\sqrt{\operatorname{med} \left(\sum_{i=1}^{n} a_{i,\Pi_n(i)}^2 \right)} + v \max_{1 \leq i,j \leq n} \{a_{i,j}\} \right) \right) \\ &\qquad \leq \mathbb{P}\big(f(A, \Pi_n) \geq t^2 \big) + \mathbb{P}\left(\sqrt{\sum_{i=1}^{n} a_{i,\Pi_n(i)}^2} \geq \sqrt{\operatorname{med} \left(\sum_{i=1}^{n} a_{i,\Pi_n(i)}^2 \right)} + v \max_{1 \leq i,j \leq n} \{a_{i,j}\} \right), \end{aligned}$$

and (3.3.2) follows from Theorem 3.1.1 and (3.2.1) in Lemma 3.2.1.

Second step: proof of (3.3.3). Consider $C_A = \mathrm{med}\,(Z)$ so that $\mathbb{P}(Z \in A) \geq 1/2$. Thus, if $v = t$ in (3.3.2)

$$\begin{aligned}\mathbb{P}\left(Z \geq \mathrm{med}\,(Z) + t\left(\sqrt{\mathrm{med}\left(\sum_{i=1}^{n} a_{i,\Pi_n(i)}^2\right)} + (t \vee C_0) \max_{1\leq i,j\leq n}\{a_{i,j}\}\right)\right) \\ \leq \mathbb{P}\left(Z \geq \mathrm{med}\,(Z) + t\left(\sqrt{\mathrm{med}\left(\sum_{i=1}^{n} a_{i,\Pi_n(i)}^2\right)} + t \max_{1\leq i,j\leq n}\{a_{i,j}\}\right)\right) \\ \leq 4e^{-t^2/16}\end{aligned}$$

Notice that the maximum with the constant in $(t \vee C_0)$ is not necessary in the case only a control of the right-tail is wanted.

Third step: proof of (3.3.4). Consider now

$$C_A = \mathrm{med}\,(Z) - t\left(\sqrt{\mathrm{med}\left(\sum_{i=1}^{n} a_{i,\Pi_n(i)}^2\right)} + v \max_{1\leq i,j\leq n}\{a_{i,j}\}\right),$$

so that

$$\mathbb{P}\left(Z \geq C_A + t\left(\sqrt{\mathrm{med}\left(\sum_{i=1}^{n} a_{i,\Pi_n(i)}^2\right)} + v \max_{1\leq i,j\leq n}\{a_{i,j}\}\right)\right) = \mathbb{P}(Z \geq \mathrm{med}\,(Z)) \geq \frac{1}{2}.$$

Hence, on the one hand, from (3.3.2),

$$\mathbb{P}(Z \in A) \leq \frac{e^{-t^2/16}}{\left(\frac{1}{2} - 2e^{-v^2/16}\right)}.$$

Thus, if $v = C_0 = 4\sqrt{\ln(8)}$, then $\left(\frac{1}{2} - 2e^{-v^2/16}\right) = \frac{1}{4}$, and $\mathbb{P}(Z \in A) \leq 4e^{-t^2/16}$.
On the other hand, as $(t \vee C_0) \geq C_0 = v$,

$$\mathbb{P}(Z \in A) \geq \mathbb{P}\left(Z \leq \mathrm{med}\,(Z) - t\left(\sqrt{\mathrm{med}\left(\sum_{i=1}^{n} a_{i,\Pi_n(i)}^2\right)} + (t \vee C_0) \max_{1\leq i,j\leq n}\{a_{i,j}\}\right)\right),$$

which ends the proof of (3.3.4).

Fourth step: proof of (3.2.3). Both (3.3.3) and (3.3.4) lead to

$$\mathbb{P}\left(|Z - \mathrm{med}\,(Z)| > t\left(\sqrt{\mathrm{med}\left(\sum_{i=1}^{n} a_{i,\Pi_n(i)}^2\right)} + (t \vee C_0) \max_{1\leq i,j\leq n}\{a_{i,j}\}\right)\right) \leq 8e^{-t^2/16}.$$

Thus, on the one hand, if $t \geq C_0$, that is $t \vee C_0 = t$, and (3.2.3) holds.
On the other hand, if $t < C_0$,

$$\begin{aligned}\mathbb{P}\left(|Z - \mathrm{med}\,(Z)| > t\left(\sqrt{\mathrm{med}\left(\sum_{i=1}^{n} a_{i,\Pi_n(i)}^2\right)} + t \max_{1\leq i,j\leq n}\{a_{i,j}\}\right)\right) &\leq 1 \\ &\leq e^{C_0^2/16 - t^2/16} = 8e^{-t^2/16},\end{aligned}$$

which ends the proof of the Proposition by taking $x = \sqrt{t}$.

3.3.3 Proof of Proposition 3.2.2

First, better readability, let

$$M = \max_{1\leq i,j\leq n}\{a_{i,j}\} \quad \text{and} \quad V = \sqrt{2\mathbb{E}\left[\sum_{i=1}^{n} a_{i,\Pi_n(i)}^2\right]} = \sqrt{\frac{2}{n}\sum_{i,j=1}^{n} a_{i,j}^2}.$$

Then, $\sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi_n(i)}^2\right)} \leq V$ since by Markov's inequality, for all non-negative random variable X, $\text{med}\,(X) \leq 2\mathbb{E}\,[X]$. Indeed,

$$\frac{1}{2} \leq \mathbb{P}(X \geq \text{med}\,(X)) \leq \frac{\mathbb{E}\,[X]}{\text{med}\,(X)}.$$

Thus, by Proposition 3.2.1, one obtains that, for all $x > 0$,

$$\mathbb{P}\big(|Z - \text{med}\,(Z)| \geq V\sqrt{x} + Mx\big) \leq 8e^{-x/16}. \tag{3.3.7}$$

The following is based on the idea of Ledoux in [114, Proposition 1.8], and provides an upper-bound of the difference between the expectation and the median of Z.

$$\begin{aligned}
|\mathbb{E}\,[Z] - \text{med}\,(Z)| &\leq \mathbb{E}\,[|Z - \text{med}\,(Z)|] \\
&= \int_0^{+\infty} \mathbb{P}(|Z - \text{med}\,(Z)| \geq r)\,dr \\
&= \int_0^{+\infty} \mathbb{P}\big(|Z - \text{med}\,(Z)| \geq V\sqrt{x} + Mx\big)\left(\frac{V}{2\sqrt{x}} + M\right)dx \\
&\leq 8\int_0^{+\infty} e^{-x/16}\left(\frac{V}{2\sqrt{x}} + M\right)dx \\
&= 8V\int_0^{+\infty} e^{-u^2/16}du + 8M\int_0^{+\infty} e^{-x/16}dx \\
&= 16V\sqrt{\pi} + 128M.
\end{aligned}$$

Therefore, denoting $\bar{\eta} := 16V\sqrt{\pi} + 128M$, one obtains from (3.3.7) that for all $x > 0$,

$$\mathbb{P}\big(|Z - \mathbb{E}\,[Z]| \geq \bar{\eta} + V\sqrt{x} + Mx\big) \leq 8e^{-x/16}. \tag{3.3.8}$$

Now, as in [26, Corollary 2.11], introduce $h_1 : u \in \mathbb{R}^+ \mapsto 1 + u - \sqrt{1+2u}$. Then, in particular, h_1 is non-decreasing, convex, one to one function on $\mathbb{R}^+$ with inverse function $h_1^{-1} : v \in \mathbb{R}^+ \mapsto v + \sqrt{2v}$. Indeed,

$$\begin{aligned}
h_1\left(h_1^{-1}(v)\right) &= 1 + v + \sqrt{2v} - \sqrt{1 + 2v + 2\sqrt{2v}} \\
&= 1 + v + \sqrt{2v} - \sqrt{\left(1 + \sqrt{2v}\right)^2} = v,
\end{aligned}$$

and

$$\begin{aligned}
h_1^{-1}\left(h_1(u)\right) &= 1 + u - \sqrt{1+2u} + \sqrt{2 + 2u - 2\sqrt{1+2u}} \\
&= u + 1 - \sqrt{1+2u} + \sqrt{1 - 2\sqrt{1+2u} + 1 + 2u} \\
&= 1 + u - \sqrt{1+2u} + \sqrt{\left(1 - \sqrt{1+2u}\right)^2} = u.
\end{aligned}$$

Consider $\mathfrak{a}$ and $\mathfrak{c}$ defined by $\mathfrak{a} = V^2/(2M)$ and $\mathfrak{c} = 2M^2/V^2$, such that $\mathfrak{ac} = M$ and $\sqrt{2\mathfrak{a}^2\mathfrak{c}} = V$ and thus

$$V\sqrt{x} + Mx = \mathfrak{a}h_1^{-1}(\mathfrak{c}x).$$

Then, from (3.3.8),

$$\mathbb{P}\big(|Z - \mathbb{E}[Z]| \geq \bar{\eta} + \mathfrak{a}h_1^{-1}(\mathfrak{c}x)\big) \leq 8e^{-x/16},$$

where $\bar{\eta} = 16\sqrt{\pi}\sqrt{2\mathfrak{a}^2\mathfrak{c}} + 128\mathfrak{ac}$.

Let $t > 0$, and consider the two following cases.

1st case: if $t \geq \bar{\eta}$, then define $x = \frac{1}{\mathfrak{c}}h_1\left(\frac{t-\bar{\eta}}{\mathfrak{a}}\right)$ such that $t = \bar{\eta} + \mathfrak{a}h_1^{-1}(\mathfrak{c}x)$. Then,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8\exp\left(-\frac{1}{16\mathfrak{c}}h_1\left(\frac{t-\bar{\eta}}{\mathfrak{a}}\right)\right).$$

Yet, by convexity of h_1,

$$h_1\left(\frac{t-\bar{\eta}}{\mathfrak{a}}\right) \geq 2h_1\left(\frac{t}{2\mathfrak{a}}\right) - h_1\left(\frac{\bar{\eta}}{\mathfrak{a}}\right).$$

Hence,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8\exp\left(\frac{1}{16\mathfrak{c}}h_1\left(\frac{\bar{\eta}}{\mathfrak{a}}\right)\right)\exp\left(-\frac{1}{8\mathfrak{c}}h_1\left(\frac{t}{2\mathfrak{a}}\right)\right).$$

Moreover, $\bar{\eta}/\mathfrak{a} = 16\sqrt{2\pi}\sqrt{\mathfrak{c}} + 128\mathfrak{c} \leq \sqrt{2 * 16\pi * 16\mathfrak{c}} + 16\pi * 16\mathfrak{c} = h_1^{-1}(16\pi * 16\mathfrak{c})$, hence

$$\frac{1}{16\mathfrak{c}}h_1\left(\frac{\bar{\eta}}{\mathfrak{a}}\right) \leq 16\pi.$$

So finally in this case,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8e^{16\pi}\exp\left(-\frac{1}{8\mathfrak{c}}h_1\left(\frac{t}{2\mathfrak{a}}\right)\right). \tag{3.3.9}$$

2nd case: if $t < \bar{\eta}$, then $h_1\left(\frac{\bar{\eta}}{2\mathfrak{a}}\right) \geq h_1\left(\frac{t}{2\mathfrak{a}}\right)$ and

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \quad \leq \quad 1 \quad \leq \quad \exp\left(\frac{1}{8\mathfrak{c}}h_1\left(\frac{\bar{\eta}}{2\mathfrak{a}}\right)\right)\exp\left(-\frac{1}{8\mathfrak{c}}h_1\left(\frac{t}{2\mathfrak{a}}\right)\right)$$

Moreover, by the same kind of upper-bounds as in the first case, one can show that

$$\frac{1}{8\mathfrak{c}}h_1\left(\frac{\bar{\eta}}{2\mathfrak{a}}\right) \leq 8\pi,$$

and thus

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq e^{8\pi}\exp\left(-\frac{1}{8\mathfrak{c}}h_1\left(\frac{t}{2\mathfrak{a}}\right)\right). \tag{3.3.10}$$

Finally, combining (3.3.9) and (3.3.10) leads, in all cases, to

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8e^{16\pi} \exp\left(-\frac{1}{8\mathfrak{c}} h_1\left(\frac{t}{2\mathfrak{a}}\right)\right). \tag{3.3.11}$$

Now, in order to obtain the Bernstein-type inequality, let $x = \frac{2}{\mathfrak{c}} h_1\left(\frac{t}{2\mathfrak{a}}\right)$, then

$$t = 2\mathfrak{a} h_1^{-1}\left(\frac{\mathfrak{c}x}{2}\right) = \mathfrak{a}\mathfrak{c}x + \sqrt{2}\sqrt{2\mathfrak{a}^2\mathfrak{c}}\sqrt{x} = Mx + \sqrt{2}V\sqrt{x},$$

and thus for all $x > 0$,

$$\mathbb{P}\left(|Z - \mathbb{E}[Z]| \geq \sqrt{2}V\sqrt{x} + Mx\right) \leq 8e^{16\pi} \exp\left(-\frac{x}{16}\right), \tag{3.3.12}$$

which ends the proof of the Proposition.

3.3.4 Proof of Corollary 3.2.1

Consider the same notation as in both Proposition 3.2.2 and its proof. This proof follows the one of [123, Corollary 2.10]. Notice that for all $u \geq 0$,

$$h_1(u) \geq \frac{u^2}{2(1+u)}.$$

Hence, from (3.3.11) in the proof of Proposition 3.2.2, for all $t \geq 0$,

$$\begin{aligned}
\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) &\leq 8e^{16\pi} \exp\left(-\frac{1}{8\mathfrak{c}} h_1\left(\frac{t}{2\mathfrak{a}}\right)\right) \\
&\leq 8e^{16\pi} \exp\left(-\frac{t^2}{64\mathfrak{a}^2\mathfrak{c}\,(1 + t/2\mathfrak{a})}\right) \\
&= 8e^{16\pi} \exp\left(-\frac{t^2}{32\,(2\mathfrak{a}^2\mathfrak{c} + \mathfrak{a}\mathfrak{c}t)}\right) \\
&= 8e^{16\pi} \exp\left(-\frac{t^2}{32\,(V^2 + Mt)}\right).
\end{aligned}$$

which ends the proof of the Corollary.

3.3.5 Proof of Theorem 3.2.1

For a better readability, introduce $a_{i,j}^+ = a_{i,j}\mathbb{1}_{a_{i,j}\geq 0}$ (respectively $a_{i,j}^- = -a_{i,j}\mathbb{1}_{a_{i,j}<0}$), and denote $Z^+ = \sum_{i=1}^n a_{i,\Pi_n(i)}^+$ (respectively $Z^- = \sum_{i=1}^n a_{i,\Pi_n(i)}^-$). Then

$$Z = \sum_{i=1}^n a_{i,\Pi_n(i)} = Z^+ - Z^-.$$

Moreover, if v (respectively v^+ and v^-) denotes $\frac{1}{n}\sum_{i,j=1}^n a_{i,j}^2$ (respectively $\frac{1}{n}\sum_{i,j=1}^n (a_{i,j}^+)^2$ and $\frac{1}{n}\sum_{i,j=1}^n (a_{i,j}^-)^2$), then $v = v^+ + v^-$ and, from the concavity property of the square root function,

$$\sqrt{2v} \geq \sqrt{v^+} + \sqrt{v^-}.$$

Furthermore, if M^+ (respectively M^-) denotes $\max_{1\leq i,j\leq n}\{a_{i,j}^+\}$ (respectively $\max_{1\leq i,j\leq n}\{a_{i,j}^-\}$), then $2M = 2\max_{1\leq i,j\leq n}\{|a_{i,j}|\} \geq M^+ + M^-$.
Finally, applying Proposition 3.2.2 to Z^+ and Z^- which are both sums of non-negative numbers leads to

$$\begin{aligned}
&\mathbb{P}\Big(|Z - \mathbb{E}[Z]| \geq 2\sqrt{2vx} + 2Mx\Big)\\
&\quad\leq \mathbb{P}\Big(\big|Z^+ - \mathbb{E}\left[Z^+\right]\big| + \big|Z^- - \mathbb{E}\left[Z^-\right]\big| \geq 2\sqrt{v^+x} + M^+x + 2\sqrt{v^-x} + M^-x\Big)\\
&\quad\leq \mathbb{P}\Big(\big|Z^+ - \mathbb{E}\left[Z^+\right]\big| \geq 2\sqrt{v^+x} + M^+x\Big)\\
&\qquad + \mathbb{P}\Big(\big|Z^- - \mathbb{E}\left[Z^-\right]\big| \geq 2\sqrt{v^-x} + M^-x\Big)\\
&\quad\leq 16e^{16\pi}\exp\left(-\frac{x}{16}\right),
\end{aligned}$$

which ends the proof of the Theorem.

3.3.6 Proof of Corollary 3.2.2

Consider the same notation as in the proof of Theorem 3.2.1, and let $t > 0$. Let M denote the maximum $\max_{1\leq i,j\leq n}\{|a_{i,j}|\}$. On the one hand, $M^+ \leq M$ and $M^- \leq M$, and on the other hand, $v^+ \leq v$ and $v^- \leq v$. Therefore, applying Corollary 3.2.1, one obtains

$$\begin{aligned}
\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) &\leq \mathbb{P}\big(\big|Z^+ - \mathbb{E}\left[Z^+\right]\big| + \big|Z^- - \mathbb{E}\left[Z^-\right]\big| \geq t\big)\\
&\leq \mathbb{P}\big(\big|Z^+ - \mathbb{E}\left[Z^+\right]\big| \geq t/2\big) + \mathbb{P}\big(\big|Z^- - \mathbb{E}\left[Z^-\right]\big| \geq t/2\big)\\
&\leq 8e^{16\pi}\left(\exp\left(\frac{-(t/2)^2}{16\left(4v^+ + 2M^+t/2\right)}\right) + \exp\left(\frac{-(t/2)^2}{16\left(4v^- + 2M^-t/2\right)}\right)\right)\\
&\leq 16e^{16\pi}\exp\left(\frac{-t^2}{64\left(4v + Mt\right)}\right),
\end{aligned}$$

which leads to the following intermediate result

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 16e^{16\pi}\exp\left(\frac{-t^2}{64\left(4\frac{1}{n}\sum_{i,j=1}^n a_{i,j}^2 + \max_{1\leq i,j\leq n}\{|a_{i,j}|\}\,t\right)}\right). \tag{3.3.13}$$

In order to make the variance appear, consider Hoeffding's centering trick recalled in (3.1.2) and introduce

$$d_{i,j} = a_{i,j} - \frac{1}{n}\sum_{k=1}^n a_{k,j} - \frac{1}{n}\sum_{l=1}^n a_{i,l} + \frac{1}{n^2}\sum_{k,l=1}^n a_{k,l} = \frac{1}{n^2}\sum_{k,l=1}^n \left(a_{i,j} - a_{k,j} - a_{i,l} + a_{k,l}\right).$$

One may easily verify that for all i_0 and j_0, $\sum_{i=1}^n d_{i,j_0} = \sum_{j=1}^n d_{i_0,j} = 0$. Moreover,

$$\sum_{i=1}^n d_{i,\Pi_n(i)} = \sum_{i=1}^n a_{i,\Pi_n(i)} - \frac{1}{n}\sum_{i,j=1}^n a_{i,j} = Z - \mathbb{E}[Z] \quad \text{and} \quad \mathbb{E}\left[\sum_{i=1}^n d_{i,\Pi_n(i)}\right] = \frac{1}{n}\sum_{i,j=1}^n d_{i,j} = 0.$$

In particular, applying equation (3.3.13) to the permuted sum of the $d_{i,j}$'s leads to

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 16e^{16\pi}\exp\left(\frac{-t^2}{64\left(4\frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2 + \max_{1\leq i,j\leq n}\{|d_{i,j}|\}\,t\right)}\right). \tag{3.3.14}$$

Then, it is sufficient to notice that, on the one hand, from [83, Theorem 2],

$$\mathrm{Var}(Z) = (n-1)^{-1} \sum_{i,j=1}^{n} d_{i,j}^2 \geq n^{-1} \sum_{i,j=1}^{n} d_{i,j}^2,$$

and on the other hand,

$$\max_{1 \leq i,j \leq n} \{|d_{i,j}|\} \leq 4 \max_{1 \leq i,j \leq n} \{|a_{i,j}|\},$$

to end the proof of Corollary 3.2.2.

3.4 Appendix: a non-asymptotic control of the second kind error

Consider the notation from the introduction of the present chapter. Since this section focuses on the study of the second kind error rate of the test, in all the sequel, the observation is assumed to satisfy the alternative $(\mathcal{H}_1)$. Let thus P be an alternative, that is $P \neq P^1 \otimes P^2$, $n \geq 4$ and $\mathbb{X}_n = (X_i, \ldots, X_n)$ be an i.i.d. sample from distribution P. Fix α and β be two fixed values in $(0,1)$. Consider T_δ the test statistic introduced in (3.1.5), the (random) critical value $q_{1-\alpha}(\mathbb{X}_n)$ defined in (3.1.6), and the corresponding permutation test defined in (3.1.7) by

$$\Delta_\alpha(\mathbb{X}_n) = \mathbb{1}_{T_\delta(\mathbb{X}_n) > q_{1-\alpha}(\mathbb{X}_n)},$$

which precisely rejects independence when $T_\delta(\mathbb{X}_n) > q_{1-\alpha}(\mathbb{X}_n)$. Notice that this test is exactly the upper-tailed test by permutation introduced in Chapter 1.

The aim of this section is to provide different conditions on the alternative P ensuring a control of the second kind error rate by a fixed value $\beta > 0$, that is $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta$. The following steps constitute the first steps of a general study of the separation rates for the previous independence test, and is worked through in the specific case of continuous real-valued random variables in Chapter 4.
For a better readability, for all real-valued measurable function g on $\mathcal{X}^2$, denote respectively

$$\mathbb{E}_P[g] = \mathbb{E}\left[g(X_1^1, X_1^2)\right] \quad \text{and} \quad \mathbb{E}_{\perp\!\!\!\perp}[g] = \mathbb{E}\left[g(X_1^1, X_2^2)\right], \tag{3.4.1}$$

the expectations of $g(X)$ under the alternative P (meaning that $X \sim P$) and under the null hypothesis $(\mathcal{H}_0)$ (meaning that $X \sim P^1 \otimes P^2$).

Assume the following moment assumption holds, that is

$(\mathcal{A}_{Mmt,2})$ | both $\mathbb{E}_P\left[\varphi_\delta^2\right] < +\infty$ and $\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right] < +\infty$,

so that all variance and second-order moments exist. Then, the following statements hold.

1. By Chebychev's inequality, one has $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta$ as soon as Condition (3.1.8) is satisfied, that is
$$\mathbb{E}\left[T_\delta(\mathbb{X}_n)\right] \geq q_{1-\beta/2}^{\alpha} + \sqrt{\frac{2}{\beta}\,\mathrm{Var}(T_\delta(\mathbb{X}_n))}.$$

2. On the one hand,
$$\mathrm{Var}(T_\delta(\mathbb{X}_n)) \leq \frac{8}{n}\left(\mathbb{E}_P\left[\varphi_\delta^2\right] + \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right]\right), \tag{3.4.2}$$

3. On the other hand, in order to control the quantile $q^{\alpha}_{1-\beta/2}$, let us first upper bound the conditional quantile, following Hoeffding's approach based on the Cauchy-Schwarz inequality, by
$$q_{1-\alpha}(\mathbb{X}_n) \leq \sqrt{\frac{1-\alpha}{\alpha}\operatorname{Var}\left(T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\middle|\mathbb{X}_n\right)}. \tag{3.4.3}$$

4. Markov's inequality allows us to deduce the following bound for the quantile:
$$q^{\alpha}_{1-\beta/2} \leq 2\sqrt{\frac{1-\alpha}{\alpha}}\sqrt{\frac{2}{\beta}\frac{\left(\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right]+\mathbb{E}_P\left[\varphi_\delta^2\right]\right)}{n}}. \tag{3.4.4}$$

5. Finally, combining (3.1.8), (3.4.2) and (3.4.4) ensures that $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n)=0)\leq\beta$ as soon as Condition (3.1.9) is satisfied, that is
$$\mathbb{E}\left[T_\delta(\mathbb{X}_n)\right] \geq \frac{4}{\sqrt{\alpha}}\sqrt{\frac{2}{\beta}\frac{\mathbb{E}\left[\varphi_\delta(X_1^1,X_1^2)^2\right]+\mathbb{E}\left[\varphi_\delta(X_1^1,X_2^2)^2\right]}{n}}.$$

6. Furthermore, under the same assumption of non-degeneracy under $(\mathcal{H}_0)$, namely $(\mathcal{A}_{non-deg})$, as in Chapter 1, that is, for $X\sim P^1\otimes P^2$, the function defined by
$$x\in\mathcal{X}\mapsto \mathbb{E}\left[\varphi_\delta(X^1,X^2)\right]-\mathbb{E}\left[\varphi_\delta(X^1,x^2)\right]-\mathbb{E}\left[\varphi_\delta(x^1,X^2)\right]+\varphi_\delta(x^1,x^2)$$
is not $P^1\otimes P^2$-almost surely equal to zero, the control in (3.4.3) together with the law of large numbers for U-statistics (see Appendix A.3.2, Theorem A.3.1) imply the consistency of the test Δ_α under all alternatives such that $\mathbb{E}\left[T_\delta(\mathbb{X}_n)\right]>0$, that is
$$\mathbb{P}(\Delta_\alpha(\mathbb{X}_n)=1)\underset{n\to+\infty}{\longrightarrow}1. \tag{3.4.5}$$

First, notice that the alternatives here are the same as the ones in Chapter 1. Moreover, notice that the consistency of the test stated in (3.4.5) is proved under weaker assumptions here than in Chapter 1 since here, only second-order moments are necessary, instead of forth-order moments in Chapter 1.

This section is divided in six subsections, each one of them respectively proving a point stated above. The first one proves the sufficiency of Condition (3.1.8) in order to control the second kind error rate. The second, third and fourth ones provide respectively upper-bounds of the variance term, the critical value and the quantile $q^{\alpha}_{1-\beta/2}$. Finally, the fifth one provides the sufficiency of Condition (3.1.9) and the last one presents an alternative proof of the consistency of the test Δ_α.

3.4.1 A first condition ensuing from Chebychev's inequality

In this section, we prove the sufficiency of a first simple condition, derived from Chebychev's inequality in order to control the second error rate. Assume that (3.1.8) is satisfied, that is
$$\mathbb{E}\left[T_\delta(\mathbb{X}_n)\right] \geq q^{\alpha}_{1-\beta/2}+\sqrt{\frac{2}{\beta}\operatorname{Var}(T_\delta(\mathbb{X}_n))}.$$

Then,

$$\begin{aligned}
\mathbb{P}(\Delta_\alpha(\mathbb{X}_n)=0) &= \mathbb{P}(T_\delta(\mathbb{X}_n)\leq q_{1-\alpha}(\mathbb{X}_n)) \qquad (3.4.6)\\
&= \mathbb{P}\Big(\{T_\delta(\mathbb{X}_n)\leq q_{1-\alpha}(\mathbb{X}_n)\}\cap\Big\{q_{1-\alpha}(\mathbb{X}_n)\leq q^\alpha_{1-\beta/2}\Big\}\Big)\\
&\quad+\mathbb{P}\Big(\{T_\delta(\mathbb{X}_n)\leq q_{1-\alpha}(\mathbb{X}_n)\}\cap\Big\{q_{1-\alpha}(\mathbb{X}_n)> q^\alpha_{1-\beta/2}\Big\}\Big)\\
&\leq \mathbb{P}\Big(T_\delta(\mathbb{X}_n)\leq q^\alpha_{1-\beta/2}\Big)+\mathbb{P}\Big(q_{1-\alpha}(\mathbb{X}_n)> q^\alpha_{1-\beta/2}\Big)\\
&\leq \mathbb{P}\Big(T_\delta(\mathbb{X}_n)\leq q^\alpha_{1-\beta/2}\Big)+\frac{\beta}{2}, \qquad (3.4.7)
\end{aligned}$$

by definition of the quantile $q^\alpha_{1-\beta/2}$. Yet, from (3.1.8) one obtains from Chebychev's inequality that

$$\begin{aligned}
\mathbb{P}\Big(T_\delta(\mathbb{X}_n)\leq q^\alpha_{1-\beta/2}\Big) &\leq \mathbb{P}\left(T_\delta(\mathbb{X}_n)\leq \mathbb{E}\left[T_\delta(\mathbb{X}_n)\right]-\sqrt{\frac{2}{\beta}\operatorname{Var}(T_\delta(\mathbb{X}_n))}\right)\\
&\leq \mathbb{P}\left(\left|T_\delta(\mathbb{X}_n)-\mathbb{E}\left[T_\delta(\mathbb{X}_n)\right]\right|\geq\sqrt{\frac{2}{\beta}\operatorname{Var}(T_\delta(\mathbb{X}_n))}\right)\\
&\leq \frac{\beta}{2}. \qquad (3.4.8)
\end{aligned}$$

Finally, both (3.4.7) and (3.4.8) lead to the desired control $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n)=0)\leq\beta$ which ends the proof.

3.4.2 Control of the variance in the general case

To upper bound the variance term, consider the following lemma based on the Cauchy-Schwartz inequality.

Lemma 3.4.1. *Let $n\geq 4$ and $\mathbb{X}_n$ be a sample of n i.i.d. random variables with distribution P and marginals P^1 and P^2. Let T_δ be the test statistic defined in (3.1.5), and $\mathbb{E}_P[\cdot]$ and $\mathbb{E}_{\perp\!\!\!\perp}[\cdot]$ be notation introduced in (3.4.1). Then, under $(\mathcal{A}_{Mmt,2})$,*

$$\operatorname{Var}(T_\delta(\mathbb{X}_n))\leq\frac{1}{n}\left(\sqrt{\mathbb{E}_P\left[\varphi_\delta^2\right]}+2\sqrt{\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right]}\right)^2.$$

In particular, this lemma implies that

$$\operatorname{Var}(T_\delta(\mathbb{X}_n))\leq\frac{2}{n}\left(\mathbb{E}_P\left[\varphi_\delta^2\right]+4\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right]\right),$$

which directly leads to (3.4.2).

Let us now prove Lemma 3.4.1. Let $n\geq 4$ and $\mathbb{X}_n$ be an i.i.d. sample with distribution P. First notice that one can write

$$T_\delta(\mathbb{X}_n)=\frac{1}{n(n-1)}\sum_{i\neq j}\left(\varphi_\delta(X_i^1,X_i^2)-\varphi_\delta(X_i^1,X_j^2)\right).$$

In particular, one recovers that $\mathbb{E}\left[T_\delta(\mathbb{X}_n)\right]=\mathbb{E}_P[\varphi_\delta]-\mathbb{E}_{\perp\!\!\!\perp}[\varphi_\delta]$.
For a better readability, let us introduce for all $i\neq j$ in $\{1,2,\dots,n\}$,

$$Y_i=\varphi_\delta(X_i^1,X_i^2)-\mathbb{E}_P[\varphi_\delta]\quad\text{and}\quad Z_{i,j}=\varphi_\delta(X_i^1,X_j^2)-\mathbb{E}_{\perp\!\!\!\perp}[\varphi_\delta].$$

Then,

$$\mathbb{E}\left[Y_i\right] = \mathbb{E}\left[Z_{i,j}\right] = 0, \quad \text{and} \quad \left\{ \begin{array}{l} \mathbb{E}\left[Y_i^2\right] = \mathrm{Var}_P(\varphi_\delta) \leq \mathbb{E}_P\left[\varphi_\delta^2\right], \\ \mathbb{E}\left[Z_{i,j}^2\right] = \mathrm{Var}_{\perp\!\!\!\perp}(\varphi_\delta) \leq \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right]. \end{array} \right. \tag{3.4.9}$$

One can write

$$T_\delta(\mathbb{X}_n) - \mathbb{E}\left[T_\delta(\mathbb{X}_n)\right] = \frac{1}{n(n-1)} \sum_{i \neq j} \left(Y_i - Z_{i,j}\right),$$

and thus,

$$\begin{aligned} \mathrm{Var}(T_\delta(\mathbb{X}_n)) &= \mathbb{E}\left[\left(\frac{1}{n(n-1)} \sum_{i \neq j} \left(Y_i - Z_{i,j}\right) \right)^2 \right] \\ &= \frac{1}{n^2(n-1)^2} \sum_{i \neq j} \sum_{k \neq l} \mathbb{E}\left[\left(Y_i - Z_{i,j}\right)\left(Y_k - Z_{k,l}\right)\right] \\ &= A_n - 2B_n + C_n, \end{aligned}$$

with

$$A_n = \frac{1}{n^2} \sum_{i,k=1}^{n} \mathbb{E}\left[Y_i Y_k\right],$$

$$B_n = \frac{1}{n^2(n-1)} \sum_{i=1}^{n} \sum_{k \neq l} \mathbb{E}\left[Y_i Z_{k,l}\right],$$

$$C_n = \frac{1}{n^2(n-1)^2} \sum_{i \neq j} \sum_{k \neq l} \mathbb{E}\left[Z_{i,j} Z_{k,l}\right],$$

where each sum is taken for indexes contained in $\{1, 2, \ldots, n\}$. In particular, since just an upper-bound of the variance is needed, it is sufficient to write

$$\mathrm{Var}(T_\delta(\mathbb{X}_n)) \leq |A_n| + 2|B_n| + |C_n|, \tag{3.4.10}$$

and to study each term separately.

Study of A_n. Since by construction, the Y_i's are centered, and independent (as the X_i's are),

$$\begin{aligned} A_n &= \frac{1}{n^2} \left(\sum_i \mathbb{E}\left[Y_i^2\right] + \sum_{i \neq k} \mathbb{E}\left[Y_i\right] \mathbb{E}\left[Y_k\right] \right) \\ &= \frac{1}{n} \mathbb{E}\left[Y_1^2\right], \end{aligned}$$

and in particular, from (3.4.9),

$$|A_n| \leq \frac{1}{n} \mathbb{E}_P\left[\varphi_\delta^2\right]. \tag{3.4.11}$$

Study of B_n. If i, k and l are all different, using once again the independence of the X_i's and a centering argument, then $\mathbb{E}[Y_i Z_{k,l}] = \mathbb{E}[Y_i]\,\mathbb{E}[Z_{k,l}] = 0$. Thus

$$\begin{aligned} B_n &= \frac{1}{n^2(n-1)} \sum_{i \neq k} \left(\mathbb{E}[Y_i Z_{i,k}] + \mathbb{E}[Y_i Z_{k,i}] \right) \\ &= \frac{1}{n} \left(\mathbb{E}[Y_1 Z_{1,2}] + \mathbb{E}[Y_1 Z_{2,1}] \right). \end{aligned}$$

In particular, applying the Cauchy-Schwartz inequality, and from (3.4.9), one obtains

$$|B_n| \leq \frac{2}{n} \sqrt{\mathbb{E}\left[Y_1^2\right] \mathbb{E}\left[Z_{1,2}^2\right]} \leq \frac{2}{n} \sqrt{\mathbb{E}_P\left[\varphi_\delta^2\right] \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right]}. \tag{3.4.12}$$

Study of C_n. Still by an independence and a centering argument, if i, j, k and l are all different, $\mathbb{E}[Z_{i,j} Z_{k,l}] = \mathbb{E}[Z_{i,j}]\,\mathbb{E}[Z_{k,l}] = 0$. Thus, if $I_n^{[3]}$ denotes the set of triplets (i,j,k) in $\{1,\dots,n\}^3$ which are all different, one obtains

$$\begin{aligned} C_n &= \frac{1}{n^2(n-1)^2} \Bigg\{ \sum_{(i,j,k) \in I_n^{[3]}} \left(\mathbb{E}[Z_{i,j} Z_{i,k}] + 2\mathbb{E}[Z_{i,j} Z_{k,i}] + \mathbb{E}[Z_{j,i} Z_{k,i}] \right) \\ &\qquad + \sum_{i \neq j} \left(\mathbb{E}\left[Z_{i,j}^2\right] + \mathbb{E}[Z_{i,j} Z_{j,i}] \right) \Bigg\} \\ &= \frac{n-2}{n(n-1)} \left(\mathbb{E}[Z_{1,2} Z_{1,3}] + 2\mathbb{E}[Z_{1,2} Z_{3,1}] + \mathbb{E}[Z_{2,1} Z_{3,1}] \right) \\ &\quad + \frac{1}{n(n-1)} \left(\mathbb{E}\left[Z_{1,2}^2\right] + \mathbb{E}[Z_{1,2} Z_{2,1}] \right). \end{aligned}$$

In particular, applying the Cauchy-Schwartz inequality, and using (3.4.9), each expectation in the previous equation satisfies $\mathbb{E}[Z_{i,j} Z_{k,l}] \leq \mathbb{E}\left[Z_{1,2}^2\right] \leq \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right]$, and thus

$$|C_n| \leq \left(\frac{4(n-2)}{n(n-1)} + \frac{2}{n(n-1)} \right) \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right] \leq \frac{4}{n} \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right]. \tag{3.4.13}$$

Finally, combining (3.4.10), (3.4.11), (3.4.12), and (3.4.13) leads to

$$\mathrm{Var}(T_\delta(\mathbb{X}_n)) \leq \frac{1}{n} \left(\sqrt{\mathbb{E}_P\left[\varphi_\delta^2\right]} + 2\sqrt{\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right]} \right)^2,$$

which ends the proof of the Lemma.

3.4.3 Control of the critical value based on Hoeffding's approach

This section is devoted to the proof the inequality (3.4.3), namely

$$q_{1-\alpha}(\mathbb{X}_n) \leq \sqrt{\frac{1-\alpha}{\alpha} \mathrm{Var}\left(T_\delta\left(\mathbb{X}_n^{\Pi_n} \right) \middle| \mathbb{X}_n \right)}.$$

The proof of this upper-bound follows Hoeffding's approach in [84], and relies on a normalizing trick, and the Cauchy-Schwarz inequality. From now on, for a better readability, denote respectively $\mathbb{E}^*[\cdot]$ and $\mathrm{Var}^*(\cdot)$ the conditional expectation and variance given the sample $\mathbb{X}_n$.

As in Hoeffding [84], the first step is to center and normalize the permuted test statistic. Yet, by construction the permuted test statistic is automatically centered, that is $\mathbb{E}^*\big[T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\big] = 0$, as one can notice that

$$T_\delta\left(\mathbb{X}_n^{\Pi_n}\right) = \frac{1}{n-1}\left(\sum_{i=1}^n \varphi_\delta\left(X_i^1, X_{\Pi_n(i)}^2\right) - \mathbb{E}^*\left[\sum_{i=1}^n \varphi_\delta\left(X_i^1, X_{\Pi_n(i)}^2\right)\right]\right).$$

Therefore, just consider the normalizing term

$$\nu_n(\mathbb{X}_n) = \mathrm{Var}^*\big(T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\big) = \mathbb{E}^*\Big[T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)^2\Big] = \frac{1}{n!}\sum_{\pi_n\in\mathfrak{S}_n}\left(T_\delta\left(\mathbb{X}_n^{\pi_n}\right)\right)^2.$$

Two cases appear: either $\nu_n(\mathbb{X}_n) = 0$ or not.

In the first case, the nullity of the conditional variance implies that all the permutations of the test statistic are equal. Hence, for all permutation π_n of $\{1,\dots,n\}$, one has $T_\delta(\mathbb{X}_n^{\pi_n}) = T_\delta(\mathbb{X}_n)$. Since the centering term $\mathbb{E}^*\Big[\sum_{i=1}^n \varphi_\delta\left(X_i^1, X_{\Pi_n(i)}^2\right)\Big] = n^{-1}\sum_{i,j=1}^n \varphi_\delta(X_i^1, X_j^2)$ is permutation invariant, one obtains the equality of the permuted sums, that is

$$\sum_{i=1}^n \varphi_\delta\left(X_i^1, X_{\pi_n(i)}^2\right) = \sum_{i=1}^n \varphi_\delta\left(X_i^1, X_i^2\right),$$

and this for all permutation π_n. In particular, the centering term is also equal to $\sum_{i=1}^n \varphi_\delta\left(X_i^1, X_i^2\right)$. Indeed, by invariance of the sum (applied in the third equality below),

$$\begin{aligned}
\frac{1}{n}\sum_{i,j=1}^n \varphi_\delta\left(X_i^1, X_j^2\right) &= \frac{1}{n}\sum_{i,j=1}^n \varphi_\delta\left(X_i^1, X_j^2\right)\left[\frac{1}{(n-1)!}\sum_{\pi_n\in\mathfrak{S}_n}\mathbb{1}_{\pi_n(i)=j}\right] \\
&= \frac{1}{n!}\sum_{\pi_n\in\mathfrak{S}_n}\sum_{i=1}^n \varphi_\delta\left(X_i^1, X_{\pi_n(i)}^2\right)\left[\sum_{j=1}^n \mathbb{1}_{\pi_n(i)=j}\right] \\
&= \frac{1}{n!}\sum_{\pi_n\in\mathfrak{S}_n}\left(\sum_{i=1}^n \varphi_\delta\left(X_i^1, X_i^2\right)\right) \\
&= \sum_{i=1}^n \varphi_\delta\left(X_i^1, X_i^2\right).
\end{aligned}$$

Therefore, $T_\delta(\mathbb{X}_n)$ is equal to zero, and thus, so is $q_{1-\alpha}(\mathbb{X}_n)$. Finally, inequality (3.4.15) is satisfied since

$$q_{1-\alpha}(\mathbb{X}_n) = 0 \le 0 = \sqrt{\frac{1-\alpha}{\alpha}\,\mathrm{Var}\Big(T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\Big|\mathbb{X}_n\Big)}.$$

Consider now the second case, and assume $\nu_n\left(\mathbb{X}_n\right) > 0$. Let us introduce the (centered and) normalized statistic

$$T'_\delta\left(\mathbb{X}_n\right) = \frac{1}{\sqrt{\nu_n(\mathbb{X}_n)}}\left(T_\delta(\mathbb{X}_n)\right).$$

In particular, the new statistic $T'_\delta\left(\mathbb{X}_n\right)$ satisfies

$$\mathbb{E}^*\big[T'_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\big] = 0 \quad \text{and} \quad \mathrm{Var}^*\big(T'_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\big) \le 1.$$

One may moreover notice that the normalizing term $\nu_n(\mathbb{X}_n)$ is permutation invariant, that is, for all permutations π_n and $\pi_n{}'$ in $\mathfrak{S}_n$,

$$\nu_n\left(\mathbb{X}_n^{\pi_n}\right) = \nu_n\left(\mathbb{X}_n\right) = \nu_n\left(\mathbb{X}_n^{\pi_n{}'}\right).$$

In particular, since $\nu_n\left(\mathbb{X}_n\right) > 0$,

$$T_\delta\left(\mathbb{X}_n^{\pi_n}\right) \leq T_\delta\left(\mathbb{X}_n^{\pi_n{}'}\right) \quad \Leftrightarrow \quad T'_\delta\left(\mathbb{X}_n^{\pi_n}\right) \leq T'_\delta\left(\mathbb{X}_n^{\pi_n{}'}\right).$$

Therefore, as the test Δ_α depends only on the comparison of the $\{T_\delta\left(\mathbb{X}_n^{\pi_n}\right)\}_{\pi_n\in\mathfrak{S}_n}$, the test statistic T_δ can be replaced by T'_δ, and the new critical value becomes

$$q'_{1-\alpha}(\mathbb{X}_n) = T'^{(n!-\lfloor n!\alpha\rfloor)}_\delta\left(\mathbb{X}_n\right) = \frac{T^{(n!-\lfloor n!\alpha\rfloor)}_\delta\left(\mathbb{X}_n\right)}{\nu_n(\mathbb{X}_n)} = \frac{q_{1-\alpha}(\mathbb{X}_n)}{\nu_n(\mathbb{X}_n)}. \tag{3.4.14}$$

Moreover, following the proof of Theorem 2.1. of Hoeffding [84], one can show (as below) that

$$q'_{1-\alpha}(\mathbb{X}_n) \leq \sqrt{\frac{1-\alpha}{\alpha}}. \tag{3.4.15}$$

Hence, combining (3.4.15) with (3.4.14) leads straightforwardly to (3.4.3).

Finally, remains the proof of (3.4.15). There are two cases:

1st case: If $q'_{1-\alpha}(\mathbb{X}_n) \leq 0$, then (3.4.15) is satisfied.

2nd case: If $q'_{1-\alpha}(\mathbb{X}_n) > 0$, then introduce $Y = q'_{1-\alpha}(\mathbb{X}_n) - T'_\delta\left(\mathbb{X}_n^{\Pi_n}\right)$.

First, since by construction, $\mathbb{E}^*\left[T'_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\right] = 0$, one directly obtains $\mathbb{E}^*[Y] = q'_{1-\alpha}(\mathbb{X}_n)$. Hence,

$$0 < q'_{1-\alpha}(\mathbb{X}_n) = \mathbb{E}^*[Y] \leq \mathbb{E}^*[Y\mathbb{1}_{Y>0}],$$

and by the Cauchy-Schwarz inequality,

$$\left(q'_{1-\alpha}(\mathbb{X}_n)\right)^2 \quad \leq \quad \left(\mathbb{E}^*[Y\mathbb{1}_{Y>0}]\right)^2 \quad \leq \quad \mathbb{E}^*\left[Y^2\right]\mathbb{E}^*[\mathbb{1}_{Y>0}],$$

Yet, on one hand,

$$\begin{aligned}
\mathbb{E}^*\left[Y^2\right] &= \mathbb{E}^*\left[\left(q'_{1-\alpha}(\mathbb{X}_n) - T'_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\right)^2\right] \\
&= \left(q'_{1-\alpha}(\mathbb{X}_n)\right)^2 + \mathbb{E}^*\left[\left(T'_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\right)^2\right] - 2q'_{1-\alpha}(\mathbb{X}_n)\mathbb{E}^*\left[T'_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\right] \\
&= \left(q'_{1-\alpha}(\mathbb{X}_n)\right)^2 + \mathrm{Var}^*\left(T'_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\right) \\
&\leq \left(q'_{1-\alpha}(\mathbb{X}_n)\right)^2 + 1,
\end{aligned}$$

since by the normalizing initial step, $\mathrm{Var}^*\left(T'_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\right) \leq 1$.

And, on the other hand,

$$\begin{aligned}
\mathbb{E}^*[\mathbb{1}_{Y>0}] &= \mathbb{E}^*\left[\mathbb{1}_{T'_\delta\left(\mathbb{X}_n^{\Pi_n}\right) < q'_{1-\alpha}(\mathbb{X}_n)}\right] \\
&= \frac{\#\left\{\pi_n \in \mathfrak{S}_n \ ; \ T'_\delta\left(\mathbb{X}_n^{\pi_n}\right) < T'^{(n!-\lfloor n!\alpha\rfloor)}_\delta\left(\mathbb{X}_n\right)\right\}}{n!} \\
&\leq \frac{(n!-\lfloor n!\alpha\rfloor)-1}{n!} = 1 - \frac{\lfloor n!\alpha\rfloor + 1}{n!} \\
&< 1 - \frac{n!\alpha}{n!} = 1-\alpha.
\end{aligned}$$

So finally,

$$\left(q'_{1-\alpha}(\mathbb{X}_n)\right)^2 \leq (1-\alpha)\left(\left(q'_{1-\alpha}(\mathbb{X}_n)\right)^2 + 1\right),$$

which is equivalent to $\left(q'_{1-\alpha}(\mathbb{X}_n)\right)^2 \leq (1-\alpha)/\alpha$, and thus ends the proof of (3.4.15).

3.4.4 Control of the quantile of the critical value

The control of the conditional quantile allows us to upper bound its own quantile $q^{\alpha}_{1-\beta/2}$ as stated in (3.4.4), that is

$$q^{\alpha}_{1-\beta/2} \leq 2\sqrt{\frac{1-\alpha}{\alpha}}\sqrt{\frac{2}{\beta}\frac{\left(\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right] + \mathbb{E}_P\left[\varphi_\delta^2\right]\right)}{n}}.$$

Indeed, (3.4.3) ensures that

$$q_{1-\alpha}(\mathbb{X}_n) \leq \sqrt{\frac{1-\alpha}{\alpha}}\sqrt{\mathbb{E}\left[T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)^2 \middle| \mathbb{X}_n\right]},$$

and in particular, the $(1-\beta/2)$-quantile of $q_{1-\alpha}(\mathbb{X}_n)$ satisfies

$$q^{\alpha}_{1-\beta/2} \leq \sqrt{\frac{1-\alpha}{\alpha}}\sqrt{\zeta_{1-\beta/2}}, \tag{3.4.16}$$

where $\zeta_{1-\beta/2}$ is the $(1-\beta/2)$-quantile of $\mathbb{E}\left[T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)^2 \middle| \mathbb{X}_n\right]$. Yet, from Markov's inequality, for all positive x,

$$\mathbb{P}\left(\mathbb{E}\left[T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)^2 \middle| \mathbb{X}_n\right] \geq x\right) \leq \frac{\mathbb{E}\left[T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)^2\right]}{x}.$$

In particular, the choice of $x = 2\mathbb{E}\left[T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)^2\right]/\beta$ leads to the control of the quantile

$$\zeta_{1-\beta/2} \leq \frac{2\mathbb{E}\left[T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)^2\right]}{\beta}. \tag{3.4.17}$$

Moreover, noticing that one can write

$$T_\delta\left(\mathbb{X}_n^{\Pi_n}\right) = \frac{1}{n-1}\sum_{i,j=1}^{n}\left(\mathbb{1}_{\Pi_n(i)=j} - \frac{1}{n}\right)\varphi_\delta(X_i^1, X_j^2),$$

the second-order moment in (3.4.17) can be rewritten

$$\begin{aligned}
\mathbb{E}\left[T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)^2\right] &= \frac{1}{(n-1)^2}\mathbb{E}\left[\left(\sum_{i,j=1}^{n}\left(\mathbb{1}_{\Pi_n(i)=j} - \frac{1}{n}\right)\varphi_\delta(X_i^1, X_j^2)\right)^2\right] \\
&= \frac{1}{(n-1)^2}\sum_{i,j=1}^{n}\sum_{k,l=1}^{n} E_{i,j,k,l} \times \mathbb{E}\left[\varphi_\delta(X_i^1, X_j^2)\varphi_\delta(X_k^1, X_l^2)\right],
\end{aligned}$$

by independence between Π_n and $\mathbb{X}_n$, where

$$E_{i,j,k,l} = \mathbb{E}\left[\left(\mathbb{1}_{\Pi_n(i)=j} - \frac{1}{n}\right)\left(\mathbb{1}_{\Pi_n(k)=l} - \frac{1}{n}\right)\right] = \mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=l}\right] - \frac{1}{n^2}.$$

On the one hand, for all $1 \leq i, j, k, l \leq n$, the Cauchy-Schwarz inequality always ensures

$$\mathbb{E}\left[\varphi_\delta(X_i^1, X_j^2)\varphi_\delta(X_k^1, X_l^2)\right] \leq \sqrt{\mathbb{E}\left[\varphi_\delta^2(X_i^1, X_j^2)\right]\mathbb{E}\left[\varphi_\delta^2(X_k^1, X_l^2)\right]} \leq \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right] + \mathbb{E}_P\left[\varphi_\delta^2\right], \tag{3.4.18}$$

since for all $1 \leq i, j \leq n$, $\mathbb{E}\left[\varphi_\delta^2(X_i^1, X_j^2)\right] \leq \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right] + \mathbb{E}_P\left[\varphi_\delta^2\right]$.
On the other hand, remains to control the sum $(n-1)^{-2}\sum_{i,j=1}^n \sum_{k,l=1}^n E_{i,j,k,l}$. Three cases appear.

1st case: If $i \neq k$ and $j \neq l$ (occurring $[n(n-1)]^2$ times), then

$$E_{i,j,k,l} = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)}.$$

2nd case: If $[i \neq k$ and $j = l]$ or $[i = k$ and $j \neq l]$, then $E_{i,j,k,l} = 0 - 1/n^2 \leq 0$.

3rd case: If $i = k$ and $j = l$ (occurring $n(n-1)$ times), then

$$E_{i,j,k,l} = \frac{1}{n} - \frac{1}{n^2} = \frac{n-1}{n^2} \leq \frac{1}{n}.$$

Therefore,

$$\begin{aligned}
\frac{1}{(n-1)^2}\sum_{i,j=1}^n \sum_{k,l=1}^n E_{i,j,k,l} &\leq \frac{1}{(n-1)^2}\left([n(n-1)]^2 \times \frac{1}{n^2(n-1)} + n(n-1) \times \frac{1}{n}\right) \\
&\leq \frac{2}{n-1} \\
&\leq \frac{4}{n}.
\end{aligned} \tag{3.4.19}$$

Finally, both (3.4.18) and (3.4.19) imply that

$$\mathbb{E}\left[T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)^2\right] \leq \frac{4}{n}\left(\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right] + \mathbb{E}_P\left[\varphi_\delta^2\right]\right), \tag{3.4.20}$$

Therefore, combining (3.4.16), (3.4.17) and (3.4.20) ends the proof of (3.4.4).

3.4.5 A new condition ensuing from Hoeffding's approach

Back to the condition (3.1.8) derived from Chebychev's inequality, both (3.4.2) and (3.4.4) imply that

$$q_{1-\beta/2}^\alpha + \sqrt{\frac{2}{\beta}\operatorname{Var}(T_\delta(\mathbb{X}_n))} \leq \sqrt{\frac{2}{\beta}\frac{\left(\mathbb{E}_P\left[\varphi_\delta^2\right] + \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right]\right)}{n}}\left(2\sqrt{\frac{1-\alpha}{\alpha}} + \sqrt{8}\right),$$

with $2\sqrt{(1-\alpha)/\alpha} + \sqrt{8} \leq 4/\sqrt{\alpha}$, since $\sqrt{1-\alpha} + \sqrt{\alpha} \leq \sqrt{2}$. Finally, the right-hand side of condition (3.1.8) being upper bounded by

$$\frac{4}{\sqrt{\alpha}}\sqrt{\frac{2}{\beta}\frac{\left(\mathbb{E}_P\left[\varphi_\delta^2\right] + \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right]\right)}{n}},$$

which is exactly the right-hand side of (3.1.9), this ensures the sufficiency of condition 3.1.9 to control the second kind error rate by β.

3.4.6 An alternative proof of the consistency of the permutation test

Assume that both assumption $(\mathcal{A}_{Mmt,2})$ and $(\mathcal{A}_{non-deg})$ are satisfied. First, applying (3.4.3) implies that

$$\begin{aligned}\mathbb{P}(\Delta_\alpha(\mathbb{X}_n)=0) &= \mathbb{P}(T_\delta(\mathbb{X}_n)\leq q_{1-\alpha}(\mathbb{X}_n))\\ &\leq \mathbb{P}\left(T_\delta(\mathbb{X}_n)\leq \sqrt{\frac{1-\alpha}{\alpha}}\sqrt{\mathrm{Var}\left(T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\middle|\mathbb{X}_n\right)}\right),\\ &= \mathbb{P}\left(T_\delta(\mathbb{X}_n)\leq \frac{1}{\sqrt{n}}\sqrt{\frac{1-\alpha}{\alpha}}\sqrt{\mathbb{E}\left[\left(\sqrt{n}T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\right)^2\middle|\mathbb{X}_n\right]}\right). \quad (3.4.21)\end{aligned}$$

Moreover, notice that, by applying the same centering trick as in Chapter 1 and introducing for all couples $x=(x^1,x^2)$ and $y=(y^1,y^2)$ in $\mathcal{X}^2$, the kernel

$$h_{\varphi_\delta}(x,y)=\frac{1}{2}\left(\varphi_\delta(x^1,x^2)+\varphi_\delta(y^1,y^2)-\varphi_\delta(x^1,y^2)-\varphi_\delta(y^1,x^2)\right),$$

the test statistic $T_\delta(\mathbb{X}_n)$ can be written as a U-statistic of order 2, that is

$$T_\delta(\mathbb{X}_n)=\frac{1}{n(n-1)}\sum_{i\neq j}h_{\varphi_\delta}\left(X_i,X_j\right).$$

Then, since $\mathbb{E}\left[|h_{\varphi_\delta}(X,X')|\right]<+\infty$ (as both $\mathbb{E}_P\left[\varphi_\delta^2\right]$ and $\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi_\delta^2\right]$ exist), the strong law of large numbers for U-statistics, stated for instance in [85] (see Appendix A.3.2, Theorem A.3.1), directly provides that

$$T_\delta(\mathbb{X}_n)\underset{n\to+\infty}{\overset{a.s.}{\longrightarrow}}\mathbb{E}\left[h_{\varphi_\delta}(X,X')\right]=\mathbb{E}\left[T_\delta(\mathbb{X}_n)\right]>0. \quad (3.4.22)$$

Furthermore, since the test statistic $T_\delta(\mathbb{X}_n)$ is exactly the U-statistic $U_{n,h_{\varphi_\delta}}$ introduced in Chapter 1 in the *Linear case*, one obtains the almost sure convergence of the conditional second-order moment (see equation (1.7.33)) which is shown in the third step of the proof of Theorem 1.4.1, that is precisely

$$\mathbb{E}\left[\left(\sqrt{n}T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\right)^2\middle|\mathbb{X}_n\right]\underset{n\to+\infty}{\overset{a.s.}{\longrightarrow}}\sigma^2_{P^1\otimes P^2}=4\,\mathrm{Var}(\mathbb{E}\left[h_{\varphi_\delta}(X,Y)|X\right])<+\infty,$$

where X and Y are two independent random variables with distribution $P^1\otimes P^2$. Notice that the non-degeneracy assumption exactly states that $\mathrm{Var}(\mathbb{E}\left[h_{\varphi_\delta}(X,Y)|X\right])$ is finite. The interested reader could refer to Section 1.2.4 in Chapter 1 of the present thesis for more details on the non-degeneracy assumption.
Hence, thanks to the multiplicative term $n^{-1/2}$, one obtains

$$\frac{1}{\sqrt{n}}\sqrt{\frac{1-\alpha}{\alpha}}\sqrt{\mathbb{E}\left[\left(\sqrt{n}T_\delta\left(\mathbb{X}_n^{\Pi_n}\right)\right)^2\middle|\mathbb{X}_n\right]}\underset{n\to+\infty}{\overset{a.s.}{\longrightarrow}}0. \quad (3.4.23)$$

Finally, combining (3.4.22) and (3.4.23) in (3.4.21) implies that $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n)=0)$ converges to 0, which ends the proof of the consistency of Δ_α.

Chapter 4

Separation rates for independence tests based on wavelet decomposition

This chapter presents an independence testing procedure in the non-parametric density framework. First, based on a wavelet thresholding method, single coefficient tests are constructed. Their corresponding critical values are obtained from a permutation approach. In particular, each of these single tests is known to be exactly of prescribed level. Then, a multiple testing procedure based on aggregation is introduced, avoiding the delicate question of the choice of the coefficient. A non-asymptotic study of the proposed procedures is performed, providing conditions on the alternative ensuring a control of the second kind error rate by a prescribed value. This leads to an upper-bound for the uniform separation rates of the aggregated test over weak Besov bodies with respect to the $\mathbb{L}_2$-metric, which should be minimax in view of the literature. Moreover, the whole procedure is adaptive as it is entirely data-driven, and its construction does not require any knowledge on the smoothness of the alternative in order to be optimal in the minimax sense.

This chapter constitutes the theoretical part of an upcoming submission. A simulation study verifying its efficiency from a practical point of view should be performed in a short time, to complete this work before submission.

Contents

4.1 Introduction 196
4.2 Description of the permutation testing procedure 199
4.2.1 The Wavelet Setting 199
4.2.2 The single coefficient testing procedure 200
4.2.3 The aggregated testing procedure 201
4.3 Theoretical Results 202
4.3.1 First and second kind error rates of the single coefficient tests 202
4.3.2 First and second kind error rates of the aggregated test 203
4.3.3 Uniform separation rates 204
4.4 Conclusion 207
4.5 Proofs 208
4.5.1 Proof of Theorem 4.3.1 208
4.5.2 Proof of Lemma 4.5.1 209

4.5.3 Proof of Proposition 4.5.1 209
4.5.4 Proof of Theorem 4.3.3 211
4.5.5 Proof of Corollary 4.3.1 213
4.6 Appendix 215
4.6.1 Proof of the level of the single coefficient tests (4.3.1) 215
4.6.2 Choice of the coefficients in the aggregated procedure 215

4.1 Introduction

This work focuses on the independence testing problem in the non-parametric density framework. Starting point in many application fields, the independence testing problem has been largely investigated in the statistical literature. From the famous chi-square test of independence from Pearson [131, 132] to recent independence tests based on reproducing kernels introduced by Gretton et al. [66], many non-parametric independence testing procedures have been developed. Those procedures are free from the underlying distribution and thus, do not require strong assumptions on the observation. Among general non-parametric testing procedures, of particular interest are those based on the permutation principle, introduced by Fisher [54] and studied, for instance, by Pitman [141], Scheffe [161] and Hoeffding [84]. The permutation approach, also referred to as the randomization approach, is a very powerful tool to construct critical values in many general hypothesis testing frameworks and this with making very few assumptions on the underlying distribution of the observations. The main advantage of this approach compared to other resampling methods such as the bootstrap "naive" one (see [51] or the introduction of this thesis for more details), is that the resulting tests are *exactly of prescribed level*, that is they control by a prescribed level, the probability under the null hypothesis to wrongly reject it, and this for any sample size of the observation. For instance, the interested reader can refer to the books of Efron and Tibshirani [51], Pesarin and Salmaso [135], or the article of Romano [155] for some reviews on permutation tests, and [4, 36, 88, 100, 111, 112] for more recent works. In this chapter, we particularly focus on Kolmogorov-Smirnov type tests of independence such as Romano's [155] and van der Vaart and Weller's [173] ones, or those introduced in Chapter 1, for which asymptotic properties are studied.
We want here to investigate the non-asymptotic performances of such tests. Let us first recall and fix a few definitions, and introduce some notation. Let $\mathbb{X}_n = (X_1, \dots, X_n)$ be an independent and identically distributed (i.i.d.) sample from the distribution P_f with density f on a subset $\mathcal{X}$ of $\mathbb{R}^d$ with respect to the Lebesgue measure. Given $\mathbb{X}_n$, a statistical test Δ of a null hypothesis $(\mathcal{H}_0)$ against an alternative $(\mathcal{H}_1)$ is here a function of $\mathbb{X}_n$ with values in $\{0, 1\}$ ($\Delta(\mathbb{X}_n) = 1$ meaning rejecting $(\mathcal{H}_0)$). Moreover, in the present work, $(\mathcal{H}_0)$ being expressed as "the coordinates of X_1 are independent", one should keep in mind that the dimension d is at least equal to two. Classically, the first kind error rate of the test Δ is the probability under $(\mathcal{H}_0)$ of wrongly rejecting it. In particular, as mentioned above, the test is said to be exactly of level α in $(0, 1)$ if its first kind error rate is controlled by α and this for any sample size n.

Permutation tests being known to guarantee a control of their first kind error rate, this work focuses on the control of the second kind error rate, that is the probability under the alternative of wrongly accepting the null hypothesis. In several articles, such as Romano's [155] and Hoeffding's [84] ones or in Chapter 1, the second kind error rate is shown to converge to

zero as the sample size tends to infinity, under any reasonable alternative. Hence, the power tends to one and the test is said to be consistent. Yet this approach is purely asymptotic, and may not be conclusive, particularly when applying the testing methods in biology (see, for instance, Chapters 1 or 2) where the sample size is often small for economical and/or biological reasons. This is why a non-asymptotic study of the second kind error rate of the tests introduced here, is performed. It provides conditions on the alternative (depending on the sample size) ensuring the control of its corresponding second kind error rate by a prescribed value. Then, the performance of the tests is studied in terms of uniform separation rates. Informally, the uniform separation rate $\rho(\Delta_\alpha, \mathcal{F}_\nu, \beta)$ of a statistical test Δ_α, with respect to (w.r.t.) some distance $\tilde{d}$, over a functional class of densities $\mathcal{F}_\nu$ of fixed smoothness parameter ν, is defined by

$$\rho\left(\Delta_\alpha, \mathcal{F}_\nu, \beta\right) = \inf\left\{\rho > 0\ ;\ \sup_{f\in\mathcal{F}_\nu,\ \tilde{d}(f,\mathcal{H}_0)>\rho} P_f(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta\right\}.$$

Intuitively, it represents the minimum distance ρ from the null hypothesis ($\mathcal{H}_0$) that guarantees the control of the second kind error rate of Δ_α by a prescribed value β in $(0,1)$ for smooth alternatives in $\mathcal{F}_\nu$. In particular, the test is optimal in a minimax sense if it reaches up to a multiplicative constant the non-asymptotic minimax rate $\rho(\mathcal{F}_\nu, \alpha, \beta)$, introduced by Baraud in [13] which is the infimum, over all tests that are exactly of level α, of their corresponding uniform separation rates. This notion is very close to the asymptotic minimax testing theory introduced by Ingster (see [91, 52] for minimax detection of signals in the Gaussian white noise model, or [93, 92] for minimax testing in the non-parametric density model).

In the following, since, as in [57, 160, 167], our study is based on a wavelet decomposition, a natural choice for $\tilde{d}$ is the $\mathbb{L}_2$-metric and legitimate choices for the functional space are Besov bodies or a weak version of them as introduced in Section 4.3.3, or Appendix A.4.2. Yet, the minimax rates of testing over d-dimensional ($d \geq 2$) weak Besov spaces are unknown. Let us have a quick and non exhaustive overview of some classical behaviors in dimension one. The choice of the distance $\tilde{d}$ and the functional class of smoothness $\mathcal{F}_\nu$ has an important impact on the minimax rate of testing, as one may see on the following examples.

- On the one hand, the choice of the metric $\tilde{d}$ may substantially influence the rates. For instance, over one-dimensional Hölder classes $H_{1,\nu,q}$ of smoothness parameter ν w.r.t. the $\mathbb{L}_q$-norm, the minimax rates obtained by Ingster [95, I section 2.4] in the non-parametric problem of testing uniformity are:
  - $n^{-2\nu/(4\nu+1)}$ if $\tilde{d}$ is the $\mathbb{L}_2$-metric,
  - $(n/\ln(n))^{-\nu/(2\nu+1)}$ if $\tilde{d}$ is the $\mathbb{L}_\infty$-metric (see also [117] for equivalent rates in the Gaussian white noise model with the classical equivalence $\varepsilon \asymp n^{-1/2}$).

  In particular, the minimax rates w.r.t. the $\mathbb{L}_2$-metric are faster than the ones obtained w.r.t. the $\mathbb{L}_\infty$-metric. The interested reader could also refer to the work of Butucea and Tribouley [32] in which they obtain the exact same rates on more general Besov spaces, and go further studying adaptation.

- On the other hand, the choice of the smoothness class may also considerably influence the rates. For instance, if $\tilde{d}$ is fixed to be the $\mathbb{L}_2$-metric, as in the testing frameworks of homogeneity in [57] or of the two-sample problem in [58], the minimax separation rates are:

- $n^{-2\nu/(4\nu+1)}$ over Besov bodies $\mathcal{B}^{\nu}_{q,\infty}(R)$ of smoothness parameter ν w.r.t. both the $\mathbb{L}_q$ and $\mathbb{L}_\infty$-norms, and radius $R > 0$ (see [57, 58] for more details),
- $(n/\ln(n))^{-\nu/(2\nu+1)}$ over weak Besov bodies $\mathcal{W}_\nu(R')$ of smoothness parameter ν, and radius $R' > 0$.

Notice that the rates obtained over weak Besov spaces are slower than the ones obtained on classical Besov spaces and seem more similar to the ones w.r.t. the $\mathbb{L}_\infty$-metric. This comes from the high irregularity of functions in weak Besov bodies which implies that testing is as hard as estimating on such functional classes, as noticed in [57]. In particular, the difference between the minimax rates (in $\mathbb{L}_2$-metric) in either estimating (see [106, 107, 45]) or testing (see [32, 57, 58]) frameworks which appears on classical Besov spaces (see, for instance, [95, I section 2.5]), disappears on weak Besov bodies. Indeed, the rate of testing over weak Besov bodies is the same as the one obtained in minimax estimation over the same functional spaces, and is closely related to the maxiset approach (see, for instance, [108, 152, 146]). Finally, one may also notice that, in [57, 58] for instance, a cost for adaptivity in $\ln(\ln(n))$ may appear in minimax rates of testing over Besov bodies, as in [167, 97], which is generally not the case for rates over weak Besov spaces.

In higher dimension $d \geq 2$, similar phenomena occur. Of main importance here, stands out the work of Ingster [94] in which are studied minimax rates of independence testing from an asymptotic distinguishability approach w.r.t. $\mathbb{L}_p$-metrics for $1 \leq p \leq +\infty$ over d-dimensional Hölder balls $H_{d,\nu,q}(R'')$ of smoothness parameter ν w.r.t. the $\mathbb{L}_q$-norm, and with radius $R'' > 0$. He obtains in particular the following rates:

- $n^{-2\nu/(4\nu+d)}$ w.r.t. the $\mathbb{L}_2$-metric,
- $(n/\ln(n))^{-\nu/(2\nu+d)}$ w.r.t. the $\mathbb{L}_\infty$-metric (see also [115] for equivalent rates in the Gaussian white noise model).

First notice that these results are coherent with the results in dimension one since the rates coincide when replacing d by 1. Furthermore, note the following.

- Once again, the rates w.r.t. the $\mathbb{L}_2$-metric are faster than the ones in $\mathbb{L}_\infty$-metric.
- Moreover, as in dimension one, the minimax rates of testing w.r.t. the $\mathbb{L}_2$-metric on classical Besov spaces $\mathcal{B}^{\nu}_{q,\infty}(R)$ obtained, for instance, by Gayraud and Tribouley in [60] (which are the same as Ingster's [94] ones above) are faster than the minimax estimation rates on Besov spaces (see, for instance, [43, Theorem 4[1]]).
- Finally, when comparing with the estimation results on weak Besov bodies $\mathcal{W}_\nu$, the rates obtained in terms of maxisets (see, for instance, Autin et al.'s article [9] based on thresholding methods) seem similar to Ingster's minimax rates stated above w.r.t. the $\mathbb{L}_\infty$-metric.

Hence, based on these remarks, it seems reasonable to conjecture that the minimax rates of testing w.r.t. the $\mathbb{L}_2$-metric over d-dimensional weak Besov bodies of smoothness parameter ν are of order $(n/\ln(n))^{-\nu/(2\nu+d)}$. In the following, the dimension is fixed to $d = 2$.

[1] with $p = \pi = 2$, $\sigma = 0$, and thus $\varepsilon > 0$.

The present work is organized as follows. Section 4.2 is devoted to the description of the testing procedure. Based on the wavelet decomposition of the difference between the density and the product of its marginals, we first introduce single coefficient tests, and then improve the testing procedure by aggregating those tests. The performance of these tests is studied in Section 4.3 where non-asymptotic conditions on the alternatives ensuring a control of the second kind error rate are given in Theorems 4.3.1 and 4.3.3. Such conditions allow us to obtain an upper-bound of the uniform separation rate over weak Besov bodies with respect to the $\mathbb{L}_2$-metric of the aggregated testing procedure (see Corollary 4.3.1). In particular, the conjectured minimax rate given above is recovered, up to a constant. After a brief conclusion in Section 4.4, the proofs are detailed in Section 4.5.

4.2 Description of the permutation testing procedure

In the present density framework, consider the set of all densities with respect to the Lebesgue measure on a compact subset $\mathcal{X}$ of $\mathbb{R}^2$. Without any loss of generality, one may assume that $\mathcal{X}$ is the unit cube $[0,1]^2$. Consider $f : [0,1]^2 \to \mathbb{R}$ such a density function, and denote by f_1 and f_2 its marginals. Then, observing an i.i.d. sample $\mathbb{X}_n = (X_1, \dots, X_n)$, where for all $1 \leq i \leq n$, $X_i = (X_i^1, X_i^2)$ is a two-dimensional random vector with density f on $[0,1]^2$, we aim at testing

$$(\mathcal{H}_0)\ f = f_1 \otimes f_2 \quad \text{against} \quad (\mathcal{H}_1)\ f \neq f_1 \otimes f_2,$$

where the (tensor) product of the marginals is defined for all (x^1, x^2) in $[0,1]^2$ by $f_1 \otimes f_2(x^1, x^2) := f_1(x^1) \times f_2(x^2)$. In particular, in all the following, the dimension is fixed to be $d = 2$. Denote by P_f the probability measure corresponding to $\mathbb{X}_n$. For simplicity, in the following, the alternative refers to either the hypothesis $(\mathcal{H}_1)$, or any density f satisfying $(\mathcal{H}_1)$. Moreover, assume that f belongs to $\mathbb{L}_\infty([0,1]^2)$, that is f satisfies $\|f\|_\infty < +\infty$. In particular, its marginals f_1 and f_2 also belong to $\mathbb{L}_\infty([0,1])$. As a consequence, f (respectively f_1 and f_2) also belongs to $\mathbb{L}_2([0,1]^2)$ (respectively $\mathbb{L}_2([0,1])$).

This work focuses on permutation tests, known for their good properties in terms of level. In particular, we consider the permutation tests introduced in Chapter 1 and adapt them to our density model. More precisely, we focus on the test statistic in the *Linear case*, which can be rewritten as

$$\frac{1}{n-1}\left(\sum_{i=1}^{n} \varphi\left(X_i^1, X_i^2\right) - \frac{1}{n} \sum_{i \neq i' \in \{1,\dots,n\}} \varphi\left(X_i^1, X_{i'}^2\right)\right).$$

The choice of a kernel φ adapted to our framework is thus fundamental. Since we want to study from a non-asymptotic point of view, the performances of our test in terms of uniform separation rates, elements of a wavelet basis seem to be a fitting choice here. For simplicity, this work considers the Haar wavelet basis of $\mathbb{L}_2([0,1]^2)$. Yet one can easily generalize it to other wavelets basis, with higher regularity for instance (see the books of Daubechies [39], Meyer [126] or the article of Cohen et al. [37]), even if boundary effects occur, since there would only be a finite number of additional father and mother wavelets to consider (see, for instance, the wavelet expansion in [145, section 3.3]).

4.2.1 The Wavelet Setting

Consider the two-dimensional isotropic Haar basis of $\mathbb{L}_2([0,1]^2)$ obtained by taking the tensor product of the one-dimensional Haar basis (see Appendix A.4.1). More precisely, for every

j in $\mathbb{N}$, and $k=(k_1,k_2)$ in $\mathcal{K}_j := \left\{0,1,\cdots,2^j-1\right\}^2$, consider the functions defined for all (x^1,x^2) in $[0,1]^2$ by

$$\varphi_0(x^1,x^2)=\phi(x^1)\phi(x^2), \quad \text{and} \quad \left\{\begin{array}{rcl} \varphi_{(1,j,k)}(x^1,x^2) & = & \phi_{j,k_1}(x^1)\psi_{j,k_2}(x^2), \\ \varphi_{(2,j,k)}(x^1,x^2) & = & \psi_{j,k_1}(x^1)\phi_{j,k_2}(x^2), \\ \varphi_{(3,j,k)}(x^1,x^2) & = & \psi_{j,k_1}(x^1)\psi_{j,k_2}(x^2), \end{array}\right.$$

where $\phi=\mathbb{1}_{[0,1)}$ and $\psi=\mathbb{1}_{[0,1/2)}-\mathbb{1}_{[1/2,1)}$ are respectively the one-dimensional Haar father and Haar mother wavelets, and

$$\Phi_{j,k}(\cdot)=2^{j/2}\Phi(2^j\cdot-k)$$

denotes the dilated/translated wavelet at scale j in $\mathbb{N}$ for Φ being either ϕ or ψ. Notice that $\mathcal{K}_j$ corresponds to the set of translations k such that for any $1\leq i\leq 3$, the intersection between the supports of the wavelets $\varphi_{(i,j,k)}$ and $[0,1)^2$ is not empty and $|\mathcal{K}_j|=2^{2j}$ where $|\cdot|$ stands for the cardinal. Then the family $\{\varphi_\lambda,\lambda\in\Lambda\}$, with

$$\Lambda=\{0\}\cup\left\{(i,j,k)\in\{1,2,3\}\times\mathbb{N}\times\mathcal{K}_j\right\},$$

is an orthonormal basis of $\mathbb{L}_2\left([0,1]^2\right)$. In particular, each function g in $\mathbb{L}_2\left([0,1]^2\right)$ can be uniquely decomposed as

$$g=\sum_{\lambda\in\Lambda}\beta_\lambda(g)\varphi_\lambda, \quad \text{where} \quad \forall\lambda\in\Lambda,\ \beta_\lambda(g)=\int_{[0,1]^2}g(x)\varphi_\lambda(x)dx. \tag{4.2.1}$$

From now on, denote $\beta_0=\beta_0(f-f_1\otimes f_2)$ the scaling coefficient, and for all λ in $\Lambda\backslash\{0\}$, $\beta_\lambda=\beta_\lambda(f-f_1\otimes f_2)$ the wavelet coefficients corresponding to the wavelet decomposition of $(f-f_1\otimes f_2)$. In particular, the null hypothesis $(\mathcal{H}_0)$ means that all the coefficients β_λ (for λ in Λ), are equal to zero, and the alternative $(\mathcal{H}_1)$ means that there exists at least one coefficient not equal to zero.
In a first step, let us construct for each wavelet function, the procedure corresponding to the test of nullity of its associated coefficient and, in a second step, let us aggregate the obtained single tests in order to construct a more powerful testing procedure of $(\mathcal{H}_0)$ against $(\mathcal{H}_1)$.

4.2.2 The single coefficient testing procedure

In all this section, consider λ a fixed index in Λ, and focus on the corresponding wavelet φ_λ. Let α in $(0,1)$ be a prescribed level. Consider the linear statistic inspired by Chapter 1, also considered in Chapter 3, defined by

$$T_\lambda(\mathbb{X}_n)=\frac{1}{n-1}\left(\sum_{i=1}^n\varphi_\lambda(X_i^1,X_i^2)-\frac{1}{n}\sum_{i,j=1}^n\varphi_\lambda(X_i^1,X_j^2)\right). \tag{4.2.2}$$

Then $T_\lambda(\mathbb{X}_n)$ is an unbiased estimator of $\int_{[0,1]^2}\varphi_\lambda(x)\left(f-f_1\otimes f_2\right)(x)dx=\beta_\lambda$, which is equal to zero under the null hypothesis $(\mathcal{H}_0)$. Based on a general thresholding idea, we construct a test that rejects the null hypothesis $(\mathcal{H}_0)$ when the test statistic $|T_\lambda(\mathbb{X}_n)|$ is too large.
Since the present work lies in a non-parametric context, the choice of the critical value is a tricky question, and is constructed as in [155] or in Chapter 1, from a permutation approach. Define for any permutation π_n of the set $\{1,2,\ldots,n\}$, the permuted sample

$$\mathbb{X}_n^{\pi_n}=(X_1^{\pi_n},\ldots,X_n^{\pi_n}) \quad \text{where for all } 1\leq i\leq n, \quad X_i^{\pi_n}=\left(X_i^1,X_{\pi_n(i)}^2\right).$$

It consists in fixing the first coordinates, and permuting only the second coordinates. Then the permuted test statistic is defined for a uniformly distributed random permutation Π_n of $\{1,\dots,n\}$, independent of the sample $\mathbb{X}_n$, by

$$T_\lambda^{\Pi_n}(\mathbb{X}_n) = T_\lambda\left(\mathbb{X}_n^{\Pi_n}\right) = \frac{1}{n-1}\left(\sum_{i=1}^n \varphi_\lambda(X_i^1, X_{\Pi_n(i)}^2) - \frac{1}{n}\sum_{i,j=1}^n \varphi_\lambda(X_i^1, X_j^2)\right). \tag{4.2.3}$$

First, one may notice that the centering term $n^{-1}\sum_{i,j=1}^n \varphi_\lambda(X_i^1, X_j^2)$ is permutation invariant. Moreover, under the null hypothesis $(\mathcal{H}_0)$, $\mathbb{X}_n^{\Pi_n}$ and $\mathbb{X}_n$ have the same distribution, and so do $|T_\lambda(\mathbb{X}_n^{\Pi_n})|$ and $|T_\lambda(\mathbb{X}_n)|$. Therefore, it is natural to consider the testing procedure that rejects $(\mathcal{H}_0)$ when $|T_\lambda(\mathbb{X}_n)| > q_{\lambda,1-\alpha}(\mathbb{X}_n)$, where for all η in $(0,1)$,

$$q_{\lambda,\eta}(\mathbb{X}_n) \text{ denotes the } \eta\text{-quantile of the conditional distribution of } |T_\lambda(\mathbb{X}_n^{\Pi_n})| \text{ given } \mathbb{X}_n. \tag{4.2.4}$$

In particular, the corresponding test function associated with the wavelet φ_λ is defined by

$$\Delta_{\lambda,\alpha}(\mathbb{X}_n) = \mathbb{1}_{|T_\lambda(\mathbb{X}_n)| > q_{\lambda,1-\alpha}(\mathbb{X}_n)}. \tag{4.2.5}$$

4.2.3 The aggregated testing procedure

The above section is devoted to single testing procedures based on one element φ_λ of the wavelet basis $\{\varphi_\lambda\}_{\lambda\in\Lambda}$ for a fixed λ in Λ. In particular, this leads to the natural and delicate question of the choice of λ. Moreover, each single coefficient test $\Delta_{\lambda,\alpha}$ focuses on a very particular type of dependence, that is "the λ-coefficient of the difference $f - f_1 \otimes f_2$ is non zero". Therefore, to avoid the choice of a particular wavelet, and to be able to detect more general forms of dependences, the idea is to consider a collection of coefficients instead of a single one, and aggregate the corresponding tests.
However, when aggregating tests, one needs to take into account the multiplicity of the tests, and therefore adapt the level of each single one of them. Several methods such as Bonferroni-type procedures have been introduced in the literature (see, for instance, [57, 59]). Yet this correction might lead to very conservative procedures, which is why we rather consider the following correction of the levels inspired by [14, 57, 58] for instance.

Consider a collection of wavelets up to some scale $\tilde{J}$ in $\mathbb{N}\cup\{+\infty\}$, that is $\{\varphi_\lambda\}_{\lambda\in\Lambda_{\tilde{J}}}$, where

$$\Lambda_{\tilde{J}} = \{0\} \cup \left\{(i,j,k) \in \{1,2,3\} \times \left\{0,1,\dots,\tilde{J}-1\right\} \times \mathcal{K}_j\right\},$$

and a collection of positive numbers $\left\{\omega_\lambda,\ \lambda\in\Lambda_{\tilde{J}}\right\}$ satisfying $\sum_{\lambda\in\Lambda_{\tilde{J}}} e^{-\omega_\lambda} \le 1$.
For each λ, let $T_\lambda(\mathbb{X}_n)$ and $T_\lambda(\mathbb{X}_n^{\Pi_n})$ be defined by (4.2.2) and (4.2.3), and consider $u_\alpha(\mathbb{X}_n)$ defined as in [58] by

$$u_\alpha(\mathbb{X}_n) = \sup\left\{u>0\ ;\ \mathbb{P}\left(\max_{\lambda\in\Lambda_{\tilde{J}}}\left\{\left|T_\lambda(\mathbb{X}_n^{\Pi_n})\right| - q_{\lambda,1-ue^{-\omega_\lambda}}(\mathbb{X}_n)\right\} > 0 \,\middle|\, \mathbb{X}_n\right) \le \alpha\right\}. \tag{4.2.6}$$

Comments.

- Informally, it is the sharpest correction of the individual levels allowing to test simultaneously all single coefficient tests with global level α. For more details on this choice of the correction $u_\alpha(\mathbb{X}_n)$ for the individual levels, the interested reader could refer to the

very recent article of Fromont et al. [59]. In particular, it provides a link between adaptive minimax testing usually based on aggregation and multiple testing procedures and proves that this choice of correction corresponds to the first step of a classical weighted min-p procedure, as defined in [47].

- Moreover, the choice of the weights $e^{-\omega_\lambda}$, $\lambda \in \Lambda_{\tilde{J}}$ and the fact that each single coefficient test is of level α implies that $u_\alpha(\mathbb{X}_n) \geq \alpha$ (see the proof of Theorem 4.3.3). Hence, replacing $u_\alpha(\mathbb{X}_n)$ by α in the following also leads to another performant (but more conservative) aggregated test.

- Notice that given an observation of $\mathbb{X}_n$, the correction $u_\alpha(\mathbb{X}_n)$ as well as the conditional quantiles $q_{\lambda,\eta}(\mathbb{X}_n)$, can be estimated by Monte Carlo methods.

Then, the aggregated testing procedure rejects the null hypothesis $(\mathcal{H}_0)$ if at least one single coefficient test at the corrected level $u_\alpha(\mathbb{X}_n)e^{-\omega_\lambda}$, namely $\Delta_{\lambda,u_\alpha(\mathbb{X}_n)e^{-\omega_\lambda}}(\mathbb{X}_n)$, rejects $(\mathcal{H}_0)$. The corresponding test function is defined by

$$\Delta_\alpha(\mathbb{X}_n) = \mathbb{1}_{\max_{\lambda\in\Lambda_{\tilde{J}}}\left\{|T_\lambda(\mathbb{X}_n)|-q_{\lambda,1-u_\alpha(\mathbb{X}_n)e^{-\omega_\lambda}}(\mathbb{X}_n)\right\}>0}. \tag{4.2.7}$$

Notice that from a theoretical aspect, as in [58] or [160], the case $\tilde{J} = +\infty$ is well-defined. Yet, it appears from the following (see Corollary 4.3.1) that $\tilde{J}$ should be chosen at least equal to $\lfloor \log_2(\sqrt{n/\ln(n)}) \rfloor$. From a practical point of view, for a matter of computational cost, $\tilde{J}$ should be as small as possible. Moreover, notice that the choice of $\tilde{J}$ does not depend on the unknown density f, or on any of its smoothness properties.

4.3 Theoretical Results

The study of the performance of these tests requires a control of their first and second kind error rates, that are respectively the probability to wrongly reject and wrongly accept the null hypothesis. It is well known that permutation tests are exactly of prescribed level.
Moreover, as we are interested here in non-asymptotic properties, instead of studying the consistency against many alternatives, as done in [155, 173] or in Chapter 1 for the same kind of permutation tests, we present, for both single coefficient and aggregated testing procedures, precise conditions on the alternatives which ensure the control of the second kind error rate by a prescribed value β in $(0, 1)$, as done in [14] in a regression model, [55] in a density model, [57] in a Poisson model, or more recently [160] in a more similar context.

4.3.1 First and second kind error rates of the single coefficient tests

By construction, the single coefficient tests are exactly of level α. This means that their first kind error rate is controlled by the prescribed level α, that is for all λ in Λ, if $f = f_1 \otimes f_2$, then, for all $n \geq 1$,

$$P_f(\Delta_{\lambda,\alpha}(\mathbb{X}_n) = 1) \leq \alpha. \tag{4.3.1}$$

This is a classical and well-known result concerning tests based on permutation of the observation. Yet, for this work to be self-contained, the proof of this result can be found in the Appendix section 4.6.1 of this chapter.

To obtain powerful tests, the second kind error rate needs to be as small as possible. Unlike the first kind error rate, the second kind error rate is not automatically controlled by construction. In particular, it depends on how far from the null hypothesis the alternative is. Thus, given a fixed value β in $(0,1)$, we want to find a non-asymptotic condition on the alternative f which guarantees a second kind error rate controlled by β, that is $P_f(\Delta_{\lambda,\alpha}(\mathbb{X}_n)=0)\leq\beta$. Intuitively, since the coefficient β_λ is equal to zero under the null hypothesis $(\mathcal{H}_0)$, the test should be more efficient in rejecting $(\mathcal{H}_0)$ for large values of $|\beta_\lambda|$. Therefore, we aim at finding a condition on the alternative of the form $|\beta_\lambda|\geq s$ for some threshold s depending on the value β, the level α, the corresponding wavelet φ_λ and the number of observations n.

Theorem 4.3.1. *Let α and β be two fixed values in $(0,1)$, $n\geq 4$ and $\mathbb{X}_n=(X_1,\dots,X_n)$ a sample of n i.i.d. random variables with density f in $\mathbb{L}_2([0,1]^2)$ satisfying $\|f\|_\infty<+\infty$. Fix λ in Λ and consider the corresponding coefficient β_λ defined in (4.2.1) and the permutation test $\Delta_{\lambda,\alpha}$ defined by (4.2.5). Let $M_f=\max\{\|f\|_\infty,\|f_1\otimes f_2\|_\infty\}$. Then, there exists two universal positive constants c and C such that $P_f(\Delta_{\lambda,\alpha}(\mathbb{X}_n)=0)\leq\beta$ as soon as*

$$|\beta_\lambda|\geq C\left\{\sqrt{\frac{2}{\beta}\frac{M_f}{n}\ln\left(\frac{c}{\alpha}\right)}+\frac{\|\varphi_\lambda\|_\infty}{n}\ln\left(\frac{c}{\alpha}\right)\right\}. \tag{4.3.2}$$

Usually in the literature (see, for instance, [58] or [160]), the control of the quantile relies on a concentration result. In our case, it directly derives from the Bernstein-type concentration inequality for randomly permuted sums obtained in Chapter 3 of this thesis. Let us recall this result stated in Theorem 3.2.1.

Theorem 4.3.2. *Let $\{a_{i,j}\}_{1\leq i,j\leq n}$ be a collection of real numbers, and Π_n be a uniformly distributed random permutation of $\{1,2,\dots,n\}$. Consider $Z:=Z(\Pi_n)=\sum_{i=1}^n a_{i,\Pi_n(i)}$. Then, there exists two universal positive constants c_0 and c_1 such that for all $x>0$,*

$$\mathbb{P}\left(|Z-\mathbb{E}[Z]|\geq 2\sqrt{2\left(\frac{1}{n}\sum_{i,j=1}^n a_{i,j}^2\right)x}+2\max_{1\leq i,j\leq n}\{|a_{i,j}|\}\,x\right)\leq c_0\exp(-c_1x). \tag{4.3.3}$$

This result directly provides a (conditional) concentration inequality for the permuted test statistic. Indeed, if $Z_n(\mathbb{X}_n)$ denotes the permuted sum $\sum_{i=1}^n\varphi_\lambda(X_i^1,X_{\Pi_n(i)}^2)$, then the centering term $n^{-1}\sum_{i,j=1}^n\varphi_\lambda(X_i^1,X_j^2)$ in (4.2.3) is nothing else but the conditional expectation $\mathbb{E}[Z_n(\mathbb{X}_n)|\mathbb{X}_n]$. One can therefore write

$$T_\lambda(\mathbb{X}_n^{\Pi_n})=\frac{1}{n-1}\left(Z_n(\mathbb{X}_n)-\mathbb{E}[Z_n(\mathbb{X}_n)|\mathbb{X}_n]\right),$$

and thus directly apply Theorem 4.3.2 with respect to the conditional probability given the sample $\mathbb{X}_n$.

4.3.2 First and second kind error rates of the aggregated test

On the one hand, as shown in [14, 55, 57, 160] in other testing frameworks, the choice of the corrected individual levels automatically implies the control of the first kind error rate of the aggregated test. Indeed, $u_\alpha(\mathbb{X}_n)$ is the greatest corrected level allowing a control of the aggregated test. Intuitively, the reason to take the maximal corrected level comes from

the fact that, the larger the level of a single test, the more often that test rejects the null hypothesis, and thus, the less conservative the procedure is.

On the other hand, as in the case of the single coefficient tests, we introduce some condition on the alternative f that guarantees a control of the second kind error rate of the aggregated procedure. Keeping in mind that we want to study the uniform separation rate w.r.t. the $\mathbb{L}_2$-metric of the aggregated test, we aim at expressing this condition directly on the $\mathbb{L}_2$-distance between f and $f_1 \otimes f_2$. Based on the Pythagorean Theorem and Theorem 4.3.1, the following statement holds.

Theorem 4.3.3. *Let α and β be two fixed values in $(0,1)$, $n \geq 4$ and $\mathbb{X}_n$ a sample of n i.i.d. random variables with density f in $\mathbb{L}_2([0,1]^2)$ satisfying $\|f\|_\infty < +\infty$. Let $\Delta_\alpha(\mathbb{X}_n)$ be the test defined by (4.2.7) for some $\tilde{J} \geq 0$ and some collection of positive numbers $\{\omega_\lambda,\ \lambda \in \Lambda_{\tilde{J}}\}$ such that $\sum_{\lambda\in\Lambda_{\tilde{J}}} e^{-\omega_\lambda} \leq 1$. Then the following properties are satisfied.*

(i) The test is exactly of level α, that is if $f = f_1 \otimes f_2$, $P_f(\Delta_\alpha(\mathbb{X}_n) = 1) \leq \alpha$.

(ii) For all subset L of $\Lambda_{\tilde{J}}$, denote by S_L the subspace of $\mathbb{L}_2([0,1]^2)$ generated by $\{\varphi_\lambda, \lambda \in L\}$, and by D_L and $\mathbf{\Pi}_L$ the dimension and the orthogonal projection onto S_L respectively. Then, there exists some positive constant $C(\alpha,\beta)$ such that $P_f(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta$ as soon as

$$\|f - f_1\otimes f_2\|_2^2 \geq \inf_{L\subset\Lambda_{\tilde{J}}} \left\{ \|(f - f_1\otimes f_2) - \mathbf{\Pi}_L(f - f_1\otimes f_2)\|_2^2 \right. \tag{4.3.4}$$
$$\left. + C(\alpha,\beta)\left[\frac{M_f}{n}\left(D_L + \sum_{\lambda\in L}\omega_\lambda\right) + \frac{2^{2J_L}}{n^2}\left(D_L + 2\sum_{\lambda\in L}\omega_\lambda + \sum_{\lambda\in L}\omega_\lambda^2\right)\right]\right\},$$

where $J_L = \max\{j \geq 0\ ;\ \exists i,k \text{ such that } (i,j,k) \in L\}$ denotes the maximum scaling level of the subset L, and M_f denotes $\max\{\|f\|_\infty, \|f_1\otimes f_2\|_\infty\}$.

Comments.

- The right-hand side of (4.3.4) is a bias-variance decomposition, the bias term taking into account the distance between the true difference $f - f_1\otimes f_2$ and its projection onto the subspace generated by the wavelets of the collection L, and the variance term containing all the contributions of each individual coefficient of the collection L.
- Furthermore, the infimum leads to adaptivity. In particular, if one can find some subset L of $\Lambda_{\tilde{J}}$, for which the inequality (4.3.4) is satisfied (without the infimum), then the second kind error rate is automatically controlled by β. This is very useful to upper bound the uniform separation rate of the aggregated test, as done in the next section.

4.3.3 Uniform separation rates

In this section, given a value β in $(0,1)$, the performance of the aggregated test is studied in terms of uniform separation rates with respect to the $\mathbb{L}_2$-metric over smoothness classes of functions $\mathcal{F}_\nu$, that are defined by

$$\rho(\Delta_\alpha, \mathcal{F}_\nu, \beta) = \inf\left\{\rho > 0\ ;\ \sup_{f\in\mathcal{F}_\nu,\ \|f - f_1\otimes f_2\|_2 > \rho} P_f(\Delta_\alpha = 0) \leq \beta\right\}. \tag{4.3.5}$$

A natural choice for functional spaces $\mathcal{F}_\nu$ are Besov bodies since they can be simply characterized in terms of the wavelet coefficients (see Appendix A.4.2). Inspired by the definition in [9] of *strong Besov bodies* and *local weak Besov spaces* in dimension $d \geq 2$, let us introduce the following definitions.

- For all smoothness parameter $\delta > 0$ and all radius $R > 0$, let us introduce the (strong) Besov body $\mathcal{B}^\delta_{2,\infty}(R)$ defined by

$$\mathcal{B}^\delta_{2,\infty}(R) = \left\{ g = \beta_0\varphi_0 + \sum_j \sum_{i,k} \beta_{(i,j,k)}\varphi_{(i,j,k)} \ ; \right. \\ \left. \beta_0^2 \leq R^2, \ \sup_{J\geq 0} \left\{ 2^{2J\delta} \sum_{j\geq J} \sum_{i,k} \beta^2_{(i,j,k)} \right\} \leq R^2. \right\} \tag{4.3.6}$$

  First notice that if the wavelet functions are regular enough, then the Besov bodies introduced above coincide with the classical ones. Moreover, one can immediately see the following inclusion for all $R > 0$,

$$\forall \delta \geq \nu, \quad \mathcal{B}^\delta_{2,\infty}(R) \subset \mathcal{B}^\nu_{2,\infty}(R). \tag{4.3.7}$$

- For all smoothness parameter $\gamma > 0$ and all radius $R' > 0$, let us introduce the d-dimensional weak Besov body $\mathcal{W}_\gamma(R')$ defined by

$$\mathcal{W}_\gamma(R') = \left\{ g = \sum_{\lambda\in\Lambda} \beta_\lambda \varphi_\lambda \ ; \ \sup_{t>0} \left\{ t^{\frac{2d}{2\gamma+d}} \sum_{\lambda\in\Lambda} \mathbb{1}_{|\beta_\lambda|>t} \right\} \leq {R'}^{\frac{2d}{2\gamma+d}} \right\}, \tag{4.3.8}$$

  with $d = 2$ in the present work.

On the one hand, Besov spaces provide ideal classes of functions when looking at the smoothness of the alternatives, whereas on the other hand, weak Besov spaces provide ideal classes of functions when looking at the sparsity of the wavelet decomposed signal (see, for instance, Rivoirard's [152] article, or Appendix A.4.2). Moreover, as stated in [9], Besov bodies and their weak versions are closely related. More precisely, one has the following property: there exists $R' > 0$ such that

$$\mathcal{B}^\gamma_{2,\infty}(R) \subset \mathcal{W}_\gamma(R'),$$

where the radius R' depends on the radius R and the smoothness parameter γ. Notably, combining this inclusion with the one in (4.3.7) shows that if $\delta \geq \gamma$, then $\mathcal{B}^\delta_{2,\infty}(R) \subset \mathcal{W}_\gamma(R')$.

Furthermore, remains the choice of the weights collection. The condition $\sum_{\lambda\in\Lambda_{\tilde{j}}} e^{-\omega_\lambda} \leq 1$ leads to a natural choice of the weights ω_λ such that one recovers a classical convergent series, as the classical p-series $\sum_{j\geq 1} j^{-2}$ for instance, and thus to set

$$\omega_0 = \ln(2) \quad \text{and} \quad \forall (i,j,k) \in \Lambda, \ \omega_{(i,j,k)} = 2\left(\ln(j+1) + \ln(\pi) + j\ln(2)\right). \tag{4.3.9}$$

The arguments leading to this choice are detailed in the Appendix section 4.6.2 of this chapter.

Then one has the following corollary of Theorem 4.3.3.

Corollary 4.3.1. *With the same notation as in Theorem 4.3.3, consider* $\tilde{J} = \left\lfloor \log_2 \left(\sqrt{n/\ln(n)} \right) \right\rfloor$ *and the family of weights introduced in* (4.3.9)*. Then for all parameters* $\delta > 0$ *and* $\gamma > 0$ *such that* $\delta \geq \gamma/(\gamma+1)$*, and all positive constants* R*,* R' *and* R''*, if*

$$\mathcal{BW}_{\delta,\gamma,\infty}(R,R',R'') = \left\{ f \in \mathbb{L}_2([0,1]^2) \; ; \; (f - f_1 \otimes f_2) \in \mathcal{B}^{\delta}_{2,\infty}(R) \cap \mathcal{W}_{\gamma}(R') \text{ and } \|f\|_{\infty} \leq R'' \right\}, \tag{4.3.10}$$

then there exists some constant $C(\alpha,\beta,\delta,\gamma,R,R',R'')$ *such that*

$$\rho\left(\Delta_{\alpha}, \mathcal{BW}_{\delta,\gamma,\infty}(R,R',R''), \beta\right) \leq C(\alpha,\beta,\delta,\gamma,R,R',R'') \left(\frac{n}{\ln(n)} \right)^{\frac{-\gamma}{2\gamma+2}}, \tag{4.3.11}$$

where the uniform separation rate ρ *is defined by* (4.3.5)*.*

Comments.

- First, notice that the smoothness assumption only concerns the difference $f - f_1 \otimes f_2$, and not necessarily the density f which can be very irregular. The idea of the proof is to construct a model L, exploiting the smoothness properties of the functional class $\mathcal{BW}_{\delta,\gamma,\infty}(R,R',R'')$ in (4.3.10), which allows us to upper bound the right-hand side of (4.3.4) in Theorem 4.3.3 by the desired rate $(n/\ln(n))^{-\gamma/(2\gamma+2)}$ in (4.3.11). In particular, both smoothness properties from $\mathcal{B}^{\delta}_{2,\infty}(R)$ and $\mathcal{W}_{\gamma}(R')$ are important to control the bias term

$$\begin{aligned} \|(f - f_1 \otimes f_2) - \mathbf{\Pi}_L(f - f_1 \otimes f_2)\|_2^2 = {} & \|(f - f_1 \otimes f_2) - \mathbf{\Pi}_{\Lambda_{\tilde{J}}}(f - f_1 \otimes f_2)\|_2^2 \\ & + \|\mathbf{\Pi}_{\Lambda_{\tilde{J}}}(f - f_1 \otimes f_2) - \mathbf{\Pi}_L(f - f_1 \otimes f_2)\|_2^2. \end{aligned}$$

  On the one hand, the (strong) Besov part $\mathcal{B}^{\delta}_{2,\infty}(R)$ in $\mathcal{BW}_{\delta,\gamma,\infty}(R,R',R'')$ ensures a fast decrease of the coefficients and thus allow the control of the first part of the bias consisted of all coefficients at scale $j \geq \tilde{J}$. On the other hand, the second part of the bias, containing all the coefficients of the collection $\Lambda_{\tilde{J}}$ which are not in the model L is controlled thanks to the sparsity properties of the weak Besov part $\mathcal{W}_{\gamma}(R')$ of $\mathcal{BW}_{\delta,\gamma,\infty}(R,R',R'')$. In particular, when comparing both obtained orders, the lower-bounding condition on δ (namely $\delta \geq \gamma/(\gamma+1)$) is required for $\lfloor \log_2(\sqrt{n/\ln(n)}) \rfloor$ to be a suitable choice for $\tilde{J}$. Similar conditions appear in dimension one, for instance, in [57], with slight differences (up to a factor two), probably due to the dimension.

- About this lower-bounding condition on δ, when comparing carefully with Autin et al.'s [9] results in a threshold estimation context, the maxiset in dimension $d = 2$ for the rate obtained in (4.3.11) is $\mathcal{B}^{\frac{\gamma}{\gamma+1}}_{2,\infty} \cap \mathcal{W}_{\gamma}$. Hence, the lower-bounding condition $\delta \geq \gamma/(\gamma+1)$ together with (4.3.7) ensure that the domain defined in (4.3.10) is included in the maxiset of [9, Theorem 3] and the result in Corollary 4.3.1 is coherent.

- Furthermore, when comparing with Ingster's [94] rate w.r.t. the $\mathbb{L}_2$-metric over Hölder spaces with smoothness parameter δ (or with the quadratic rates over Besov spaces obtained in [60]), that is $n^{-2\delta/(4\delta+2)}$, one may notice that the intersection with the weak Besov body allows an improvement of the rate for $\delta < \gamma/2$ thanks to the sparsity properties. The exact same cut appears in dimension one in [57, 58, 160]. Hence, this

result is attractive for parameters satisfying $\gamma/(\gamma+1) \leq \delta < \gamma/2$, which requires that $\gamma > 1$. A similar condition, more precisely $\gamma > 1/2$, appears also in dimension one in a similar setting [57]. The transition from $1/2$ to 1 is probably due to the dimension.

- Moreover, notice that, since the dimension here is $d = 2$, the separation rate obtained in the corollary is equal to $(n/\ln(n))^{-\gamma/(2\gamma+d)}$. This result is coherent compared to results in dimension one (as in [57, 58, 160]). Moreover, as mentioned above, it is exactly the same rate as the one obtained by Autin et al. in [9] in the estimation context from a maxiset approach on similar classes of functions. It is also similar to the rates obtained by Ingster [96] in the minimax independence testing framework w.r.t. the $\mathbb{L}_\infty$-metric over Hölder balls with smoothness parameter δ. In particular, it achieves the minimax rate conjectured in the introduction of this chapter.

- Finally, notice that the testing procedure (and in particular the choice of the maximal scale $\tilde{J}$) is fully data-driven and does not require any knowledge on the smoothness parameters δ or γ. In particular, up to the fact that the lower-bounds are not known, but only conjectured, the testing procedure is hence said to be adaptive in the minimax sense.

4.4 Conclusion

In the present work, we introduced non-parametric procedures to test the independence hypothesis in the density framework. The proposed critical values are obtained from a permutation approach which guarantees good non-asymptotic properties for the tests in terms of level. First, we introduce single tests based on linear test statistics based on wavelet functions. Then, we construct a more general and powerful procedure by aggregating a collection of such single coefficient tests. Finally, we study the performance of the aggregated test in terms of uniform separation rates, providing a minimum distance (w.r.t. the $\mathbb{L}_2$-metric) from the null hypothesis guaranteeing the control of the second kind error rate for smooth enough alternatives. The smoothness functional spaces considered here are weak Besov bodies.

Two main open questions remain. The first one is to compute lower-bounds for the minimax rates of testing over d-dimensional weak Besov bodies with respect to the $\mathbb{L}_2$-metric for d larger than or equal to two. In particular, it would allow us to justify that our testing procedure is optimal in the minimax sense. The second remaining question is to generalize this method in order to construct a testing procedure which reaches, up to a constant, the minimax rates of testing over classical Besov (or more particularly Hölder) spaces, for which the minimax rates are known. Notice that in the literature, such asymptotically minimax-optimal independence testing procedures have been constructed (see Ingster [96] and Yodé [180], and [181] with adaptivity). Yet in all these articles, the performances of the tests are studied from an asymptotic point of view, whereas we are interested in non-asymptotic properties. In particular, on the one hand, they do not control non-asymptotically the levels. Moreover, on the other hand, they may be very conservative. However, one may notice that all these testing procedures are based on U-statistics (which is also the case for minimax-optimal testing procedures in dimension one as in [32, 57, 58]). Hence, permuted U-statistics should allow us to construct minimax-optimal testing procedures over Besov spaces, and will be the subject of a further work.

4.5 Proofs

In all the proofs, for a better readability, set

$$\mathbb{E}_f[\varphi] = \int_{[0,1]^2} \varphi(x) f(x) dx \quad \text{and} \quad \mathbb{E}_{f_1 \otimes f_2}[\varphi] = \int_{[0,1]^2} \varphi(x) f_1 \otimes f_2(x) dx.$$

4.5.1 Proof of Theorem 4.3.1

Since the critical value is random, we follow the idea of Fromont et al. in [58] (as in Chapter 3), and introduce $q^{\alpha}_{1-\beta/2}$ the $(1-\beta/2)$-quantile of the conditional quantile $q_{\lambda,1-\alpha}(\mathbb{X}_n)$ defined in (4.2.4). The proof is based on the following results.

The first lemma, based on Chebychev's inequality, provides a general condition on the coefficient β_λ ensuring the control of the second kind error rate of the test by the prescribed value β.

Lemma 4.5.1. *Consider the same notation as in Theorem 4.3.1 and let $q^{\alpha}_{1-\beta/2}$ be the $(1-\beta/2)$-quantile of the conditional quantile $q_{\lambda,1-\alpha}(\mathbb{X}_n)$.*
Then, by Chebychev's inequality, $P_f(\Delta_{\lambda,\alpha}(\mathbb{X}_n) = 0) \leq \beta$ as soon as

$$|\beta_\lambda| \geq q^{\alpha}_{1-\beta/2} + \sqrt{\frac{2}{\beta} \operatorname{Var}(T_\lambda(\mathbb{X}_n))}. \tag{4.5.1}$$

Since the terms in the right-hand side of (4.5.1) are difficult to compute, the following general proposition allows us to control each of them separately. Notice that this result is true for more general functions φ_λ possibly depending on some parameter λ.

Proposition 4.5.1. *Consider the same notation as in Theorem 4.3.1 and let $q^{\alpha}_{1-\beta/2}$ be the $(1-\beta/2)$-quantile of the conditional quantile $q_{\lambda,1-\alpha}(\mathbb{X}_n)$. Then, on the one hand,*

$$\operatorname{Var}(T_\lambda(\mathbb{X}_n)) \leq \frac{2}{n}\left(\mathbb{E}_f\left[\varphi_\lambda^2\right] + 4\mathbb{E}_{f_1\otimes f_2}\left[\varphi_\lambda^2\right]\right), \tag{4.5.2}$$

and on the other hand, there exists two universal positive constants C and c_0 such that

$$q^{\alpha}_{1-\beta/2} \leq C\left\{\sqrt{\frac{2}{\beta}\ln\left(\frac{c_0}{\alpha}\right)}\left(\frac{\sqrt{\mathbb{E}_f\left[\varphi_\lambda^2\right]}}{n} + \frac{\sqrt{\mathbb{E}_{f_1\otimes f_2}\left[\varphi_\lambda^2\right]}}{\sqrt{n}}\right) + \frac{\|\varphi_\lambda\|_\infty}{n}\ln\left(\frac{c_0}{\alpha}\right)\right\}. \tag{4.5.3}$$

Furthermore, by definition of the translated/dilated wavelet φ_λ here, $\|\varphi_\lambda\|_2^2 = 1$ and thus both $\mathbb{E}_f\left[\varphi_\lambda^2\right]$ and $\mathbb{E}_{f_1\otimes f_2}\left[\varphi_\lambda^2\right]$ are smaller than M_f. Indeed, for g being either f or $f_1 \otimes f_2$ both uniformly bounded on $[0,1]^2$, $\mathbb{E}_g\left[\varphi_\lambda^2\right] = \int_{[0,1]^2} \varphi_\lambda(x) g(x) dx \leq \|g\|_\infty \|\varphi_\lambda\|_2^2 = \|g\|_\infty$. Therefore, Proposition 4.5.1 implies that the right-hand side of (4.5.1) is upper bounded by

$$C\left\{\sqrt{\frac{2}{\beta}\frac{M_f}{n}\left(\ln\left(\frac{c_0}{\alpha}\right)+1\right)} + \frac{\|\varphi_\lambda\|_\infty}{n}\ln\left(\frac{c_0}{\alpha}\right)\right\},$$

itself upper bounded by the right-hand side of (4.3.2) for $c = c_0 \exp(1)$ universal positive constant.
Finally, if (4.3.2) is satisfied, then so is (4.5.1) and therefore, Lemma 4.5.1 ensures that $P_f(\Delta_{\lambda,\alpha}(\mathbb{X}_n) = 0) \leq \beta$ which ends the proof of Theorem 4.3.1.

4.5.2 Proof of Lemma 4.5.1

This proof is very close in spirit to the proof of Condition 3.1.8 in Chapter 3. The main difference comes from the absolute values. In particular, notice that here, $q_{\lambda,1-\alpha}(\mathbb{X}_n)$ is the $(1-\alpha)$-quantile of $\mathcal{L}\left(\left|T_\lambda(\mathbb{X}_n^{\Pi_n})\right|\middle|\mathbb{X}_n\right)$ and not $\mathcal{L}\left(T_\lambda(\mathbb{X}_n^{\Pi_n})\middle|\mathbb{X}_n\right)$.
Let us first prove that condition (4.5.1) guarantees a control of the second kind error rate by the fixed value β. In particular, since $T_\lambda(\mathbb{X}_n)$ is an unbiased estimator of β_λ, (4.5.1) is equivalent to

$$q^\alpha_{1-\beta/2} \leq |\mathbb{E}\left[T_\lambda(\mathbb{X}_n)\right]| - \sqrt{\frac{2}{\beta}\operatorname{Var}(T_\lambda(\mathbb{X}_n))},$$

which implies that

$$q^\alpha_{1-\beta/2} \leq \mathbb{E}\left[|T_\lambda(\mathbb{X}_n)|\right] - \sqrt{\frac{2}{\beta}\operatorname{Var}(|T_\lambda(\mathbb{X}_n)|)}.$$

Therefore, separating the cases $\left\{q_{\lambda,1-\alpha}(\mathbb{X}_n) \leq q^\alpha_{1-\beta/2}\right\}$ or $\left\{q_{\lambda,1-\alpha}(\mathbb{X}_n) > q^\alpha_{1-\beta/2}\right\}$, one obtains

$$\begin{aligned}
P_f(\Delta_{\lambda,\alpha}(\mathbb{X}_n)=0) &= \mathbb{P}(|T_\lambda(\mathbb{X}_n)| \leq q_{\lambda,1-\alpha}(\mathbb{X}_n)) \\
&\leq \mathbb{P}\left(|T_\lambda(\mathbb{X}_n)| \leq q^\alpha_{1-\beta/2}\right) + \mathbb{P}\left(q_{\lambda,1-\alpha}(\mathbb{X}_n) > q^\alpha_{1-\beta/2}\right) \\
&\leq \mathbb{P}\left(|T_\lambda(\mathbb{X}_n)| - \mathbb{E}\left[|T_\lambda(\mathbb{X}_n)|\right] \leq -\sqrt{\frac{2}{\beta}\operatorname{Var}(|T_\lambda(\mathbb{X}_n)|)}\right) + \frac{\beta}{2} \\
&\leq \mathbb{P}\left(\left||T_\lambda(\mathbb{X}_n)| - \mathbb{E}\left[|T_\lambda(\mathbb{X}_n)|\right]\right| \geq \sqrt{\frac{2}{\beta}\operatorname{Var}(|T_\lambda(X_n)|)}\right) + \frac{\beta}{2} \\
&\leq \beta,
\end{aligned}$$

by Chebychev's inequality which ends the proof of Lemma 4.5.1.

4.5.3 Proof of Proposition 4.5.1

Control of the variance. To upper bound the variance term, we apply Lemma 3.4.1 in Section 3.4.2 which directly provides that

$$\begin{aligned}
\operatorname{Var}(T_\lambda(\mathbb{X}_n)) &\leq \frac{1}{n}\left(\sqrt{\mathbb{E}_f\left[\varphi_\lambda^2\right]} + 2\sqrt{\mathbb{E}_{f_1\otimes f_2}\left[\varphi_\lambda^2\right]}\right)^2 \\
&\leq \frac{2}{n}\left(\mathbb{E}_f\left[\varphi_\lambda^2\right] + 4\mathbb{E}_{f_1\otimes f_2}\left[\varphi_\lambda^2\right]\right),
\end{aligned}$$

which is exactly equation (4.5.2).

Control of the quantile. The proof of (4.5.3) is divided into two steps. The first step consists in controlling the conditional quantile $q_{\lambda,1-\alpha}(\mathbb{X}_n)$ and the second step provides an upper-bound for $q^\alpha_{1-\beta/2}$.

1st step. Let us first give an upper-bound for the conditional quantile $q_{\lambda,1-\alpha}(\mathbb{X}_n)$. Applying Theorem 4.3.2, there exist some universal positive constants C' and c_0 such that

$$q_{\lambda,1-\alpha}(\mathbb{X}_n) \leq \frac{C'}{n-1}\left\{\sqrt{\frac{1}{n}\sum_{i,j=1}^n \varphi_\lambda^2(X_i^1, X_j^2)}\sqrt{\ln\left(\frac{c_0}{\alpha}\right)} + \|\varphi_\lambda\|_\infty \ln\left(\frac{c_0}{\alpha}\right)\right\}. \tag{4.5.4}$$

Indeed, introduce $Z(\mathbb{X}_n) = \sum_{i=1}^n \varphi_\lambda(X_i^1, X_{\Pi_n(i)}^2)$. Then, notice that

$$T_\lambda^{\Pi_n}(\mathbb{X}_n) = \frac{1}{n-1}\left(Z(\mathbb{X}_n) - \mathbb{E}\left[Z(\mathbb{X}_n)|\mathbb{X}_n\right]\right). \tag{4.5.5}$$

Therefore, applying Theorem 4.3.2 to the conditional probability given $\mathbb{X}_n$, one obtains that there exist universal positive constants c_0 and c_1 such that, for all $x > 0$,

$$\mathbb{P}\left(\left|Z(\mathbb{X}_n) - \mathbb{E}\left[Z(\mathbb{X}_n)|\mathbb{X}_n\right]\right| \geq 2\sqrt{2\left(\frac{1}{n}\sum_{i,j=1}^n \varphi_\lambda^2(X_i^1, X_j^2)\right)x} + 2\|\varphi_\lambda\|_\infty x \middle| \mathbb{X}_n\right) \leq c_0 \exp\left(-c_1 x\right).$$

In particular, from (4.5.5), one obtains

$$\mathbb{P}\left(\left|T_\lambda(\mathbb{X}_n^{\Pi_n})\right| \geq \frac{2}{n-1}\left(\sqrt{2\left(\frac{1}{n}\sum_{i,j=1}^n \varphi_\lambda^2(X_i^1, X_j^2)\right)x} + \|\varphi_\lambda\|_\infty x\right) \middle| \mathbb{X}_n\right) \leq c_0 \exp\left(-c_1 x\right).$$

Yet, by definition of the quantile, $q_{\lambda,1-\alpha}(\mathbb{X}_n)$ is the smallest u such that

$$\mathbb{P}\left(\left|T_\lambda(\mathbb{X}_n^{\Pi_n})\right| \geq u \middle| \mathbb{X}_n\right) \leq \alpha.$$

Thus taking x such that $c_0 \exp\left(-c_1 x\right) = \alpha$, that is $x = c_1^{-1} \ln\left(c_0/\alpha\right)$, one obtains (4.5.4) with $C' = 2\max\left\{\sqrt{2/c_1}, 1/c_1\right\}$ which is a universal positive constant.

2nd step. Let us now control the quantile $q_{1-\beta/2}^\alpha$. Since (4.5.4) is always true, by definition of $q_{1-\beta/2}^\alpha$, one has that $q_{1-\beta/2}^\alpha$ is upper bounded by the $(1-\beta/2)$-quantile of the right-hand side of (4.5.4). Yet, the only randomness left in the right-hand side of (4.5.4) comes from the randomness of $\frac{1}{n}\sum_{i,j=1}^n \varphi_\lambda^2(X_i^1, X_j^2)$, and thus it is sufficient to control its $(1-\beta/2)$-quantile.

Besides, applying Markov's inequality, one obtains for all $x > 0$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i,j=1}^n \varphi_\lambda^2(X_i^1, X_j^2) \geq x\right) \leq \frac{\mathbb{E}\left[\frac{1}{n}\sum_{i,j=1}^n \varphi_\lambda^2(X_i^1, X_j^2)\right]}{x},$$

with $\mathbb{E}\left[\frac{1}{n}\sum_{i,j=1}^n \varphi_\lambda^2(X_i^1, X_j^2)\right] = \mathbb{E}_f\left[\varphi_\lambda^2\right] + (n-1)\mathbb{E}_{f_1\otimes f_2}\left[\varphi_\lambda^2\right]$, and thus, taking

$$x = \frac{2}{\beta}\left(\mathbb{E}_f\left[\varphi_\lambda^2\right] + (n-1)\mathbb{E}_{f_1\otimes f_2}\left[\varphi_\lambda^2\right]\right),$$

one has that the $(1-\beta/2)$-quantile of $\frac{1}{n}\sum_{i,j=1}^n \varphi_\lambda^2(X_i^1, X_j^2)$ is upper bounded by x, and thus, the $(1-\beta/2)$-quantile of $\sqrt{\frac{1}{n}\sum_{i,j=1}^n \varphi_\lambda^2(X_i^1, X_j^2)}$ is itself upper bounded by

$$\sqrt{\frac{2}{\beta}}\left(\sqrt{\mathbb{E}_f\left[\varphi_\lambda^2\right]} + \sqrt{n}\sqrt{\mathbb{E}_{f_1\otimes f_2}\left[\varphi_\lambda^2\right]}\right).$$

Finally,

$$q^{\alpha}_{1-\beta/2} \leq \frac{2C'}{n}\left\{\sqrt{\frac{2}{\beta}}\left(\sqrt{\mathbb{E}_f\left[\varphi_\lambda^2\right]}+\sqrt{n}\sqrt{\mathbb{E}_{f_1\otimes f_2}\left[\varphi_\lambda^2\right]}\right)\sqrt{\ln\left(\frac{c_0}{\alpha}\right)}+\|\varphi_\lambda\|_\infty \ln\left(\frac{c_0}{\alpha}\right)\right\}.$$

which is exactly (4.5.3) for any constant $C \geq 2C'$.

4.5.4 Proof of Theorem 4.3.3

Proof of *(i)* Consider a fixed level α in $(0,1)$. Let $n \geq 1$ and $\mathbb{X}_n$ a sample of n i.i.d. random variables with density $f = f_1 \otimes f_2$ with respect to the Lebesgue measure and consider the correction of the level $u_\alpha(\mathbb{X}_n)$ defined in (4.2.6) and the test Δ_α defined by (4.2.7) for $\tilde{J}$ in $\mathbb{N}\cup\{+\infty\}$ and a collection of weights $\{\omega_\lambda, \lambda \in \Lambda_{\tilde{J}}\}$.
First notice that under the null hypothesis $(\mathcal{H}_0)$, for all permutation π_n of $\{1,\dots,n\}$, the (deterministically) permuted sample $\mathbb{X}_n^{\pi_n}$ has the same distribution as $\mathbb{X}_n$, so $|T_\lambda(\mathbb{X}_n^{\pi_n})|$ and $|T_\lambda(\mathbb{X}_n)|$ have the same distribution.
Moreover, if Π_n is a uniformly distributed random permutation independent of $\mathbb{X}_n$, so is $\Pi_n \circ \pi_n$. Hence the conditional distributions $\mathcal{L}\left(|T_\lambda^{\Pi_n}(\mathbb{X}_n^{\pi_n})|\middle|\mathbb{X}_n^{\pi_n}\right)$ and $\mathcal{L}\left(|T_\lambda^{\Pi_n}(\mathbb{X}_n)|\middle|\mathbb{X}_n\right)$ are equal which leads to

$$q_{\lambda,1-\alpha}(\mathbb{X}_n^{\pi_n}) = q_{\lambda,1-\alpha}(\mathbb{X}_n) \quad \text{and} \quad u_\alpha(\mathbb{X}_n^{\Pi_n}) = u_\alpha(\mathbb{X}_n).$$

Therefore, if $\mathfrak{S}_n$ denotes the set of permutations of $\{1,\dots,n\}$, then

$$\begin{aligned}
&\mathbb{P}\big(\Delta_\alpha(\mathbb{X}_n)=1\big)\\
&\quad= \mathbb{P}\left(\max_{\lambda\in\Lambda_{\tilde{J}}}\left\{|T_\lambda(\mathbb{X}_n)| - q_{\lambda,1-u_\alpha(\mathbb{X}_n)e^{-\omega_\lambda}}(\mathbb{X}_n)\right\}>0\right)\\
&\quad= \frac{1}{n!}\sum_{\pi_n\in\mathfrak{S}_n}\mathbb{P}\left(\max_{\lambda\in\Lambda_{\tilde{J}}}\left\{|T_\lambda(\mathbb{X}_n^{\pi_n})| - q_{\lambda,1-u_\alpha(\mathbb{X}_n^{\pi_n})e^{-\omega_\lambda}}(\mathbb{X}_n^{\pi_n})\right\}>0\right)\\
&\quad= \sum_{\pi_n\in\mathfrak{S}_n}\mathbb{P}\left(\max_{\lambda\in\Lambda_{\tilde{J}}}\left\{\left|T_\lambda(\mathbb{X}_n^{\Pi_n})\right| - q_{\lambda,1-u_\alpha(\mathbb{X}_n)e^{-\omega_\lambda}}(\mathbb{X}_n)\right\}>0\middle|\Pi_n=\pi_n\right)\mathbb{P}(\Pi_n=\pi_n)\\
&\quad= \mathbb{P}\left(\max_{\lambda\in\Lambda_{\tilde{J}}}\left\{\left|T_\lambda(\mathbb{X}_n^{\Pi_n})\right| - q_{\lambda,1-u_\alpha(\mathbb{X}_n)e^{-\omega_\lambda}}(\mathbb{X}_n)\right\}>0\right)\\
&\quad= \mathbb{E}\left[\mathbb{P}\left(\max_{\lambda\in\Lambda_{\tilde{J}}}\left\{\left|T_\lambda^{\Pi_n}(\mathbb{X}_n)\right| - q_{\lambda,1-u_\alpha(\mathbb{X}_n)e^{-\omega_\lambda}}(\mathbb{X}_n)\right\}>0\middle|\mathbb{X}_n\right)\right]\\
&\quad\leq \alpha,
\end{aligned}$$

by definition of $u_\alpha(\mathbb{X}_n)$, which ends the proof.

Proof of *(ii)* First, one can easily show that $u_\alpha(\mathbb{X}_n) \geq \alpha$. Indeed, applying Bonferroni's inequality,

$$\begin{aligned}
\mathbb{P}\left(\max_{\lambda\in\Lambda_{\tilde{J}}}\left\{\left|T_\lambda(\mathbb{X}_n^{\Pi_n})\right| - q_{\lambda,1-\alpha e^{-\omega_\lambda}}(\mathbb{X}_n)\right\}>0\middle|\mathbb{X}_n\right) &\leq \sum_{\lambda\in\Lambda_{\tilde{J}}}\mathbb{P}\left(\left|T_\lambda(\mathbb{X}_n^{\Pi_n})\right| > q_{\lambda,1-\alpha e^{-\omega_\lambda}}(\mathbb{X}_n)\middle|\mathbb{X}_n\right)\\
&\leq \sum_{\lambda\in\Lambda_{\tilde{J}}}\alpha e^{-\omega_\lambda} \leq \alpha,
\end{aligned}$$

by definition of the conditional quantiles and the choice of the weights ω_λ. In particular, setting $\alpha_\lambda = \alpha e^{-\omega_\lambda}$, one obtains that for all λ in L, $q_{\lambda,1-u_\alpha(\mathbb{X}_n)e^{-\omega_\lambda}}(\mathbb{X}_n) \leq q_{\lambda,1-\alpha_\lambda}(\mathbb{X}_n)$, and

$$\begin{aligned}
\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) &= \mathbb{P}\left(\min_{\lambda\in\Lambda_{\tilde{J}}} \left\{|T_\lambda(\mathbb{X}_n)| - q_{\lambda,1-u_\alpha(\mathbb{X}_n)e^{-\omega_\lambda}}(\mathbb{X}_n)\right\} \leq 0\right) \\
&\leq \mathbb{P}\left(\min_{\lambda\in\Lambda_{\tilde{J}}} \{|T_\lambda(\mathbb{X}_n)| - q_{\lambda,1-\alpha_\lambda}(\mathbb{X}_n)\} \leq 0\right) \\
&\leq \min_{\lambda\in\Lambda_{\tilde{J}}} \mathbb{P}(|T_\lambda(\mathbb{X}_n)| \leq q_{\lambda,1-\alpha_\lambda}(\mathbb{X}_n)) \\
&\leq \min_{\lambda\in\Lambda_{\tilde{J}}} \mathbb{P}(\Delta_{\lambda,\alpha_\lambda}(\mathbb{X}_n) = 0)\,.
\end{aligned}$$

Hence, if at least one of the single coefficient test $\Delta_{\lambda,\alpha_\lambda}(\mathbb{X}_n)$ with corrected level α_λ controls the second kind error rate by β, then $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta$. Therefore, applying Theorem 4.3.1, it is sufficient to show that condition (4.3.4) implies that there exists λ in the collection satisfying condition (4.3.2) for the corrected level α_λ, that is

$$|\beta_\lambda| \geq C\left\{\sqrt{\frac{2}{\beta}\frac{M_f}{n}\ln\left(\frac{c}{\alpha_\lambda}\right)} + \frac{\|\varphi_\lambda\|_\infty}{n}\ln\left(\frac{c}{\alpha_\lambda}\right)\right\}, \tag{4.5.6}$$

where C and c are the universal constants from Theorem 4.3.1, to end the proof of Theorem 4.3.3.

Since $\|\mathbf{\Pi}_L(f - f_1\otimes f_2)\|_2^2 = \sum_{\lambda\in L}\beta_\lambda^2$, let us first bound from above for any λ in L, the square of the right-hand side of (4.5.6), denoted by RHS_λ here for better readability.

Consider $\lambda = (i,j,k)$ in L. Then $\|\varphi_\lambda\|_\infty = 2^j$, and since $\ln(c/\alpha_\lambda) = \ln(c/\alpha) + \omega_\lambda$,

$$\begin{aligned}
{RHS_\lambda}^2 &\leq 2C^2\left\{\frac{2}{\beta}\frac{M_f}{n}\left(\ln\left(\frac{c}{\alpha}\right) + \omega_\lambda\right) + \frac{2^{2j}}{n^2}\left(\ln\left(\frac{c}{\alpha}\right) + \omega_\lambda\right)^2\right\} \\
&\leq C(\alpha,\beta)\left\{\frac{M_f}{n}(1+\omega_\lambda) + \frac{2^{2j}}{n^2}(1+\omega_\lambda)^2\right\},
\end{aligned} \tag{4.5.7}$$

with $C(\alpha,\beta) = 2C^2\ln(c/\alpha)\max\{2/\beta, \ln(c/\alpha)\}$.

Hence, if (4.3.4) is satisfied for this constant, then, applying the Pythagorean Theorem, there exists some subset L of $\Lambda_{\tilde{J}}$ satisfying

$$\|\mathbf{\Pi}_L(f - f_1\otimes f_2)\|_2^2 \geq C(\alpha,\beta)\left\{\frac{M_f}{n}\left(D_L + \sum_{\lambda\in L}\omega_\lambda\right) + \frac{2^{2J_L}}{n^2}\left(D_L + 2\sum_{\lambda\in L}\omega_\lambda + \sum_{\lambda\in L}\omega_\lambda^2\right)\right\},$$

which is equivalent to

$$\sum_{\lambda\in L}\beta_\lambda^2 \geq \sum_{\lambda\in L} C(\alpha,\beta)\left\{\frac{M_f}{n}(1+\omega_\lambda) + \frac{2^{2j}}{n^2}(1+\omega_\lambda)^2\right\}.$$

In particular, there exists $\lambda = (i,j,k) \in L$ such that

$$\beta_\lambda^2 \geq C(\alpha,\beta)\left\{\frac{M_f}{n}(1+\omega_\lambda) + \frac{2^{2j}}{n^2}(1+\omega_\lambda)^2\right\},$$

which combined with (4.5.7) leads to condition (4.5.6). Therefore, Theorem 4.3.1 allows us to conclude that $\mathbb{P}(\Delta_{\lambda,\alpha_\lambda}(\mathbb{X}_n) = 0) \leq \beta$ for that particular λ, and therefore, $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta$ which ends the proof of Theorem 4.3.3.

4.5.5 Proof of Corollary 4.3.1

In all this proof, $C(\alpha, \beta, \ldots)$ is a constant depending on the parameters α, $\beta, \ldots$ which may vary from one line to another.
Let f be in $\mathcal{BW}_{\delta,\gamma,\infty}(R, R', R'')$. From Theorem 4.3.3, it is sufficient to find some $L \subset \Lambda_{\tilde{J}}$ and some constant depending on all the parameters α, β, δ, γ, R, R' and R'' such that

$$\|(f-f_1\otimes f_2)-\mathbf{\Pi}_L(f-f_1\otimes f_2)\|_2^2+C(\alpha,\beta,\|f\|_\infty)\left[\frac{1}{n}\left(D_L+\sum_{\lambda\in L}\omega_\lambda\right)+\frac{2^{2j_L}}{n^2}\left(D_L+\sum_{\lambda\in L}\omega_\lambda^2\right)\right]$$
$$\leq C(\alpha,\beta,\delta,\gamma,R,R',R'')\left(\frac{n}{\ln(n)}\right)^{\frac{-\gamma}{\gamma+1}}. \quad (4.5.8)$$

First notice that for this particular choice of weights, since for all $L \subset \Lambda_{\tilde{J}}$, and for all (i,j,k) in L,

$$\omega_{(i,j,k)} = 2\left(\ln(j+1)+\ln(\pi)+j\ln(2)\right) \leq 4\ln(2)(j+1) \leq 4\ln(2)(j_L+1),$$

where j_L is the largest scaling level of the subset L as defined in Theorem 4.3.3. In particular, for any $L \subset \Lambda_{\tilde{J}}$, the left-hand side of (4.5.8) is upper bounded by

$$\|(f-f_1\otimes f_2)-\mathbf{\Pi}_L(f-f_1\otimes f_2)\|_2^2+C(\alpha,\beta,R'')\left[\frac{j_L+1}{n}D_L+\frac{2^{2j_L}(j_L+1)^2}{n^2}D_L\right],$$

since $\|f\|_\infty \leq R''$. Yet, $j_L \leq \tilde{J} \leq \log_2\left(\sqrt{n/\ln(n)}\right)$, which implies that, for all $n \geq 4$,

$$2^{2j_L}\frac{(j_L+1)}{n} \leq \frac{n}{\ln(n)}\frac{1}{n}\left(\frac{\ln(n)-\ln(\ln(n))}{2\ln(2)}+1\right) \leq \frac{1}{\ln(2)}.$$

Hence, for any $L \subset \Lambda_{\tilde{J}}$, the left-hand side of (4.5.8) is thus upper bounded by

$$\|(f-f_1\otimes f_2)-\mathbf{\Pi}_L(f-f_1\otimes f_2)\|_2^2+C(\alpha,\beta,R'')\left(\frac{j_L+1}{n}D_L\right). \quad (4.5.9)$$

Let us now construct some $L \subset \Lambda_{\tilde{J}}$ such that the above bound (4.5.9) is itself upper bounded by some constant times the rate $(n/\ln(n))^{-\gamma/(\gamma+1)}$.

Consider $D \leq |\Lambda_{\tilde{J}}|+1 = 2^{2\tilde{J}}$ an integer to be determined later. Let L be the subset of $\Lambda_{\tilde{J}}$ consisted of the indices of the D largest coefficients among the coefficients $\{|\beta_\lambda|\}_{\lambda\in\Lambda_{\tilde{J}}}$. In particular, $D_L = D$. Moreover, Pythagorean Theorem leads to the following decomposition:

$$\|(f-f_1\otimes f_2)-\mathbf{\Pi}_L(f-f_1\otimes f_2)\|_2^2 = Q_1+Q_2,$$

where

$$\begin{cases} Q_1 &= \|(f-f_1\otimes f_2)-\mathbf{\Pi}_{\Lambda_{\tilde{J}}}(f-f_1\otimes f_2)\|_2^2, \\ Q_1 &= \|\mathbf{\Pi}_{\Lambda_{\tilde{J}}}(f-f_1\otimes f_2)-\mathbf{\Pi}_L(f-f_1\otimes f_2)\|_2^2. \end{cases}$$

On the one hand, since $(f-f_1\otimes f_2)$ belongs to $\mathcal{B}^\delta_{2,\infty}(R)$,

$$Q_1 = \sum_{j\geq\tilde{J}}\sum_{i,k}\beta^2_{(i,j,k)} \leq R^2 2^{-2\tilde{J}\delta} = R^2\left(\frac{n}{\ln(n)}\right)^{-\delta},$$

which comes from the choice of $\tilde{J}$.
On the other hand, since $(f - f_1 \otimes f_2)$ belongs to $\mathcal{W}_\gamma(R')$, for all $t > 0$,

$$\sum_{j\geq 0}\sum_{i,k} \mathbb{1}_{|\beta_{(i,j,k)}|>t} \leq (R')^{\frac{2}{\gamma+1}} t^{-\frac{2}{\gamma+1}}.$$

Thus, choosing t such that $(R')^{\frac{2}{\gamma+1}} t^{-\frac{2}{\gamma+1}} = D$, that is $t = (R'D)^{-\frac{\gamma+1}{2}}$, implies that the number of coefficients larger than t is less than D, and in particular, all coefficients with index in $\Lambda_{\tilde{J}} \backslash L$ are less than or equal to t. Therefore,

$$\begin{aligned}
Q_2 &= \sum_{\lambda\in\Lambda_{\tilde{J}}\backslash L} \beta_\lambda^2 \leq \sum_{\lambda\in\Lambda} \beta_\lambda^2 \mathbb{1}_{|\beta_\lambda|\leq t} \\
&\leq \sum_{\lambda\in\Lambda} \beta_\lambda^2 \sum_{l=0}^{+\infty} \mathbb{1}_{2^{-(l+1)}t<|\beta_\lambda|\leq 2^{-l}t} \\
&\leq t^2 \sum_{l=0}^{+\infty} 2^{-2l} \sum_{\lambda\in\Lambda} \mathbb{1}_{|\beta_\lambda|>2^{-l}t/2} \\
&\leq t^2 \sum_{l=0}^{+\infty} 2^{-2l} (R')^{\frac{2}{\gamma+1}} \left(\frac{t}{2}\right)^{\frac{-2}{\gamma+1}} 2^{l\frac{2}{\gamma+1}} \\
&\leq (2R')^{\frac{2}{\gamma+1}} t^{\frac{2\gamma}{\gamma+1}} \sum_{l=0}^{+\infty} \left(2^{\frac{-2\gamma}{\gamma+1}}\right)^l \\
&\leq C(\gamma, R') D^{-\gamma}.
\end{aligned}$$

Finally, the bias term is controlled as follows

$$\|(f - f_1 \otimes f_2) - \mathbf{\Pi}_L(f - f_1 \otimes f_2)\|_2^2 \leq C(\delta, \gamma, R, R') \left(\left(\frac{n}{\ln(n)}\right)^{-\delta} + D^{-\gamma} \right). \tag{4.5.10}$$

Hence, by combining the bounds (4.5.9) and (4.5.10) and using the fact that $j_L \leq \tilde{J}$, the left-hand side of (4.5.8) is upper bounded by

$$C(\alpha, \beta, \delta, \gamma, R, R', R'') \left(\left(\frac{n}{\ln(n)}\right)^{-\delta} + D^{-\gamma} + \frac{\tilde{J}+1}{n} D \right). \tag{4.5.11}$$

The last step consists in choosing D in order to obtain the same order (in n) for each element of the sum in (4.5.11).
Consider D to be of the form $\lfloor (n/\ln(n))^\eta \rfloor$. Then $D^{-\gamma}$ and $(J+1)D/n$ are of the same order as soon as $(n/\ln(n))^{-\gamma\eta}$ and $(n/\ln(n))^{\eta-1}$ are of the same order, that is $\eta = 1/(\gamma+1)$.
Hence, considering

$$D = \left\lfloor \left(\frac{n}{\ln(n)}\right)^{\frac{1}{1+\gamma}} \right\rfloor,$$

both $D^{-\gamma}$ and $(J+1)D/n$ are of the order of $(n/\ln(n))^{-\gamma/(\gamma+1)}$. Notice that since $\gamma > 0$, the condition $D \leq 2^{2\tilde{J}}$ is always satisfied.

Remains the term $2^{-2J\delta} = (n/\ln(n))^{-\delta} = \mathcal{O}\big((n/\ln(n))^{-\gamma/(\gamma+1)}\big)$ since the smoothness parameters are assumed to satisfy $\delta \geq \gamma/(\gamma+1)$. Thus, one obtains for this choice of D that (4.5.11) is upper bounded by

$$C(\alpha,\beta,\delta,\gamma,R,R',R'')\left(\frac{n}{\ln(n)}\right)^{\frac{-\gamma}{\gamma+1}}.$$

Finally, if $\|f - f_1\otimes f_2\|_2 \geq C(\alpha,\beta,\delta,\gamma,R,R',R'')\left(\frac{n}{\ln(n)}\right)^{\frac{-\gamma}{2\gamma+2}}$, then (4.3.4) in Theorem 4.3.3 is satisfied and $P_f(\Delta_\alpha(\mathbb{X}_n)=0) \leq \beta$, which ends the proof of Corollary 4.3.1.

4.6 Appendix

4.6.1 Proof of the level of the single coefficient tests (4.3.1)

This proof is very close in spirit with the one presented in the introduction of this thesis. Fix an index λ in $\{0\}\cup\Lambda$, and a level α in $(0,1)$. Let $n\geq 1$ and $\mathbb{X}_n$ be a sample of n i.i.d. random variables with density $f = f_1\otimes f_2$ with respect to the Lebesgue measure and consider the test $\Delta_{\lambda,\alpha}$ defined by (4.2.5). Then, as under the null hypothesis ($\mathcal{H}_0$), for all permutation π_n of $\{1,\dots,n\}$, the (deterministically) permuted sample $\mathbb{X}_n^{\pi_n}$ has the same distribution as $\mathbb{X}_n$, so $|T_\lambda(\mathbb{X}_n^{\pi_n})|$ and $|T_\lambda(\mathbb{X}_n)|$ have the same distribution.
Moreover, if Π_n is a uniformly distributed random permutation independent of $\mathbb{X}_n$, so is $\Pi_n\circ\pi_n$. Furthermore, the σ-algebra generated by $\mathbb{X}_n$ is exactly the same as the one generated by $\mathbb{X}_n^{\pi_n}$ since the application $\mathbb{X}_n \mapsto \mathbb{X}_n^{\pi_n}$ is measurable and bijective. Hence the conditional distributions $\mathcal{L}\left(|T_\lambda^{\Pi_n}(\mathbb{X}_n^{\pi_n})|\Big|\mathbb{X}_n^{\pi_n}\right)$ and $\mathcal{L}\left(|T_\lambda^{\Pi_n}(\mathbb{X}_n)|\Big|\mathbb{X}_n\right)$ are equal which leads to

$$q_{\lambda,1-\alpha}(\mathbb{X}_n^{\pi_n}) = q_{\lambda,1-\alpha}(\mathbb{X}_n).$$

Therefore, if $\mathfrak{S}_n$ denotes the set of permutations of $\{1,\dots,n\}$, then, by definition of the conditional quantile $q_{\lambda,1-\alpha}(\mathbb{X}_n)$,

$$\begin{aligned}
P_f(\Delta_{\lambda,\alpha}(\mathbb{X}_n)=1) &= \frac{1}{n!}\sum_{\pi_n\in\mathfrak{S}_n}\mathbb{P}(|T_\lambda(\mathbb{X}_n^{\pi_n})|\geq q_{\lambda,1-\alpha}(\mathbb{X}_n^{\pi_n}))\\
&= \sum_{\pi_n\in\mathfrak{S}_n}\mathbb{P}\big(|T_\lambda(\mathbb{X}_n^{\Pi_n})|\geq q_{\lambda,1-\alpha}(\mathbb{X}_n)\big|\Pi_n=\pi_n\big)\,\mathbb{P}(\Pi_n=\pi_n)\\
&= \mathbb{P}\big(|T_\lambda(\mathbb{X}_n^{\Pi_n})|\geq q_{\lambda,1-\alpha}(\mathbb{X}_n)\big)\\
&= \mathbb{E}\left[\mathbb{P}\big(|T_\lambda(\mathbb{X}_n^{\Pi_n})|\geq q_{\lambda,1-\alpha}(\mathbb{X}_n)\big|\mathbb{X}_n\big)\right]\\
&\leq \alpha,
\end{aligned}$$

which ends the proof.

4.6.2 Choice of the coefficients in the aggregated procedure

First, notice that the condition $\sum_{\lambda\in\Lambda_{\tilde{J}}} e^{-\omega_\lambda}\leq 1$ is always implied by

$$e^{-\omega_0}\leq\frac{1}{2}\quad\text{and}\quad\sum_{j\geq 0}\sum_{i=1}^{3}\sum_{k\in\mathcal{K}_j}e^{-\omega_{(i,j,k)}}\leq\frac{1}{2}. \tag{4.6.1}$$

Hence, one can set $e^{-\omega_0} = 1/2$, that is $\omega_0 = \ln(2)$.
Moreover, since

$$\sum_{j\geq 0} \frac{1}{(j+1)^2} = \frac{\pi^2}{6},$$

the second condition in (4.6.1) is equivalent to

$$\sum_{j\geq 0}\sum_{i=1}^{3}\sum_{k\in\mathcal{K}_j} e^{-\omega_{(i,j,k)}} \leq \frac{3}{\pi^2}\sum_{j\geq 0}\frac{1}{(j+1)^2}.$$

Furthermore, if one considers weights that do not depend on i and k, but only on the scaling resolution j, the initial condition becomes,

$$\forall i \in \{1,2,3\}\,,\ \forall k \in \mathcal{K}_j,\ \omega_{(i,j,k)} = \omega_j, \qquad \text{with} \qquad \sum_{j\geq 0} 3\,|\mathcal{K}_j|\,e^{-\omega_j} \leq \sum_{j\geq 0}\frac{3}{\pi^2(j+1)^2}.$$

Hence, a natural idea is to consider weights ω_j's satisfying $|\mathcal{K}_j|\,e^{-\omega_j} = (\pi(j+1))^{-2}$, that is

$$\omega_j = 2(\ln(j+1) + \ln(\pi)) + \ln\left(|\mathcal{K}_j|\right),$$

with $|\mathcal{K}_j| = 2^{2j}$, which directly provides (4.3.9).

Comment. Notice that one could have chosen any convergent series as a reference, such as geometric series for instance. Yet, as mentioned above, the larger the individual corrected levels, the less conservative the aggregated procedure. Hence, this explains the choice of p-series which have larger terms.

Conclusion and perspectives (english)

This work revisits very old procedures based on bootstrap or permutation approaches and adapts them to functionals of non-product type, and point processes as random variables. This allows us to elaborate more reliable statistical methods for the problem of synchrony detection in neuroscience. Moreover, the asymptotic behavior study of the conditional distribution of the permuted test statistic given the observation led us to a new combinatorial central limit theorem which holds under any alternative (see Theorem 1.4.1). To our knowledge, no combinatorial central limit theorem has been proved for non-deterministic and non-exchangeable variables, as it is the case under the dependence alternative, and this, even in other settings than the point process framework. As a consequence, the permutation approach is shown to reconstruct the null hypothesis, and in this sense, those results partially fulfill van der Vaart and Wellner's open question.
Moreover, we went further in a density framework, and studied the non-asymptotic properties of our permutation test of independence in the *Linear case* (that is based on U-statistics with kernels h of the form h_φ as in (0.4.3)). To do so, we proved a new Bernstein-type concentration inequality for permuted sums (see Theorem 3.2.1), which allows us to understand the non-asymptotic performance of our tests through the study of their uniform separation rate. However, many questions remain open.

The most immediate perspective would be to complete Chapter 4 with a simulation study to check the efficiency of our aggregated procedure in practice, and to investigate its limits.

Moreover, concerning the non-asymptotic study, we provide upper-bounds for the uniform separation rate of our permutation test over weak Besov bodies with respect to the $\mathbb{L}_2$-metric. Yet, the lower-bounds of the minimax separation rate on such smoothness classes are missing at this point. This opens several perspectives.

- The first one is to study these lower-bounds in order to justify that our test is optimal, and thus adaptive, in the minimax sense.
- The other ones rely on the lower-bounds for the minimax rates of independence testing over Hölder spaces with respect to the $\mathbb{L}_p$-metrics for $1 \leq p \leq \infty$ obtained by Ingster [94].
  - The minimax rates found by Ingster for $p = \infty$ are really close to the ones we obtained over weak Besov bodies, and we hope that our permutation test also matches those rates. A further study is in progress.
  - Moreover, a brief look at the literature suggests that U-statistics are required to construct minimax optimal procedures over Hölder spaces with respect to the $\mathbb{L}_2$-metric. The idea is thus to generalize our testing method to non-linear choices of U-statistics (by considering (0.4.2) based on kernels h not being of the form

h_φ) in order to achieve the minimax rates of testing provided by Ingster [94] and Yodé [180, 181].

Yet, as explained above, concentration inequalities are necessary to obtain sharp controls of the quantiles. The main perspective here is thus to extend the Bernstein-type concentration inequalities for permuted sums to permuted U-statistics. It will be the subject of a further work.

New perspectives also open up in neuroscience.

- One of the main remaining questions is the calibration of the unknown parameter δ in the delayed coincidence count φ_δ^{coinc}. In line with the Kolmogorov-Smirnov type tests of Romano or van der Vaart and Wellner, a natural idea would be to take a supremum over different kernel functions h of the form $h_{\varphi_\delta^{coinc}}$. More generally, from a theoretical point of view, it would be interesting to know whether our results from Chapter 1 can be extended to test statistics of the form $\sup_h U_{n,h}$ where the supremum is taken over reasonable classes of functions. Even in the *Linear case* $h = h_\varphi$, it would already be a generalization to not necessarily product type functions φ's. However, in neuroscience, doing this might lose valuable information which may interest neurobiologists. Indeed, δ represents the typical delay of interaction, that is informally the time a neuron needs to transmit the information to another. A more suitable idea, borrowed from Chapter 4, would be to aggregate several tests based on different values of δ taken in a grid for instance, with corrected individual levels taking the multiplicity of the tests into account, and reject independence if at least one of the single test does. In particular, the delays δ for which the independence is rejected are then available.

  A starting point for the statistical analysis of such procedure is the Bernstein-type concentration inequality obtained in Chapter 3. For each δ, it should provide sharp bounds on the corresponding quantile providing conditions on the alternative guaranteeing the control of the second kind error rates. A simulation study needs also to be done in order to validate such procedure from a practical point of view.

- Other types of function φ may also be investigated. In particular, another possible choice for φ would be to consider Sansonnet and Tuleau-Malot's statistic introduced in [160], and to combine the permutation approach with an aggregation procedure based on their choice of φ, to see if the rates match the ones they prove under restrictive distribution assumptions (namely, the Poissonian interaction model).

- Concerning the *Permutation Unitary Events method* introduced in Chapter 2, remains the theoretical validity of the multiple testing procedure. Indeed, for now, the control of the false discovery rate is only guaranteed in the very particular case of homogeneous Poisson processes, disjoint windows, and only upper-tailed (or only lower-tailed) tests, so that the single tests are independent. We have investigated the validity of the property of positive regression dependency (PRDS) on the subset of all true hypotheses in the other cases, but it does not seem to hold. Hence, even though the constructed procedure seems to perform well from a practical point of view, in order to obtain a fully theoretically justified procedure, one should rather apply other procedures, such as the step-up procedures based on different shape functions (of a c.d.f. form) considered by Blanchard and Roquain in [23], as in particular, the distribution free linear step-up procedure of Benjamini and Yekutieli [17] replacing α by $\alpha \times (\sum_{j=1}^m 1/j)^{-1}$ in Benjamini and Hochberg's procedure.

- Finally, one of the main drawbacks of the method introduced here to treat the synchrony detection issue in spike train analysis, is that it only applies for two neurons. Therefore, a first step would be to generalize this method to three or more neurons, and to be able to detect different patterns of interacting neurons. In particular, one needs to generalize the permutation approach to several spike trains in a pattern, and apply the generalized delayed coincidence count introduced by Chevallier and Laloë in [35]. Then, in the line of their work, to avoid the tricky choice of which pattern to examine, the idea is to perform another step of multiple testing, each single test representing a different pattern. Once again, the theoretical validity of such multiple testing remains to be proved.

In even more general frameworks, Gretton and his co-authors constructed some independence tests based on reproducing kernels (see, for instance, [65, 66]). Their test statistic is based on a U-statistic, and the corresponding critical values are obtained from a permutation approach in practice. It would be interesting to study their testing procedure from a non-asymptotic point of view, as in this thesis, and compare the obtained rates to ours. Yet, to do so, the generalization of the concentration inequality for permuted sums to permuted U-statistics mentioned above, needs to be accomplished.

Conclusion et perspectives (français)

Ce travail reprend des procédures de tests étudiées par Romano [155] puis van der Vaart et Wellner [173], basées sur des approches par bootstrap ou par permutation. Nous les adaptons à des fonctionnelles qui peuvent ne pas être de type produit, et aux processus ponctuels. Ceci nous permet alors d'élaborer des méthodes statistiques fiables pour répondre à la problématique de la détection de synchronisations en neurosciences. Par ailleurs, l'étude du comportement asymptotique de la loi conditionnelle de la statistique permutée sachant l'observation nous a mené à un nouveau théorème de la limite centrale combinatoire valable sous n'importe quelle alternative (voir le théorème 1.4.1). À notre connaissance, aucun théorème de ce genre n'a été démontré pour des variables aléatoires non-échangeables, comme c'est le cas ici sous l'alternative de dépendance, et ce, même dans d'autres contextes que celui des processus ponctuels. Par conséquent, nous avons montré que l'approche par permutation permet de reconstruire la loi sous $(\mathcal{H}_0)$, et, en ce sens, ces résultats répondent partiellement à la question laissée ouverte par van der Vaart et Wellner [173].
D'autre part, nous sommes allés plus loin dans le cadre de variables aléatoires à densité en étudiant les propriétés non-asymptotiques de notre test par permutation dans le *cas linéaire* (c'est-à-dire, basé sur des U-statistiques dont le noyau h est de la forme h_φ comme dans (0.1.4)). Pour cela, nous avons démontré une inégalité de concentration de type Bernstein pour les sommes permutées (voir le théorème 3.2.1), nous permettant ensuite d'en déduire les performances non-asymptotiques de notre test à travers l'étude de leur vitesse de séparation uniforme. Cependant, de nombreuses questions restent ouvertes.

La première perspective est de compléter le chapitre 4 en vérifiant l'efficacité de notre procédure de tests agrégés d'un point de vue pratique et d'en étudier les limites grâce à une étude par simulation.

Par ailleurs, lors de l'étude des performances non-asymptotiques, nous obtenons une borne supérieure de la vitesse de séparation uniforme de notre test par permutation sur des espaces de Besov faibles par rapport à la distance $\mathbb{L}_2$. Cependant, les bornes inférieures ne sont pas connues à ce jour. Cela ouvre de nouvelles perspectives.

- La première est d'étudier ces bornes inférieures afin de justifier que notre test est optimal, et donc adaptatif au sens du minimax sur de tels espaces.
- Les autres perspectives reposent sur les bornes inférieures obtenues par Ingster [94] pour les vitesses minimax des tests d'indépendance sur des espaces de Hölder par rapport aux distances $\mathbb{L}_p$ où $1 \leq p \leq \infty$.
  - Les vitesses de séparation uniformes que nous avons obtenues en distance $\mathbb{L}_2$ sur les espaces de Besov faibles sont très proches des vitesses minimax obtenues par

Ingster pour $p = +\infty$, et nous espérons que notre test par permutation atteint également ces vitesses en distance $\mathbb{L}_\infty$. Une étude complémentaire est en cours.

– Par ailleurs, un rapide coup d'œil à la littérature suggère que les U-statistiques sont nécessaires pour construire des procédures optimales au sens du minimax sur des espaces de Hölder par rapport à la distance $\mathbb{L}_2$. L'idée est donc de généraliser notre procédure de test à des choix non-linéaires de U-statistiques (en considérant (0.1.3) basée sur des noyaux h n'étant pas de la forme h_φ) afin d'atteindre les vitesses de test minimax obtenues par Ingster [94] et Yodé [180, 181]. Cependant, comme nous l'avons mentionné précédemment, les inégalités de concentration sont nécessaires pour obtenir des contrôles suffisamment fins des quantiles. La plus grande perspective ici est donc d'étendre les inégalités de concentration de type Bernsetin pour les sommes permutées à des U-statistiques permutées. Ceci sera étudié dans des travaux à venir.

De nouvelles perspectives s'ouvrent également en neurosciences.

- L'une des principales questions restantes concerne la calibration du paramètre inconnu δ dans la notion de coïncidence avec délai φ_δ^{coinc}. Dans la lignée des tests de type Kolmogorov-Smirnov de Romano ou de van der Vaart et Wellner, une idée naturelle serait de considérer un supremum sur différents noyaux h de la forme $h_{\varphi_\delta^{coinc}}$. Plus généralement, d'un point de vue théorique, il serait intéressant de savoir si les résultats présentés dans le chapitre 1 peuvent s'étendre à des statistiques de test de la forme $\sup_h U_{n,h}$ où le supremum est pris sur des classes raisonnables de fonctions. Même dans le *cas linéaire*, ce serait déjà une généralisation à des fonctions φ n'étant pas nécessairement de la forme produit. Cependant, en neurosciences, cela entrainerait la perte d'une information intéressante pour les neuroscientifiques. En effet, δ représente le délai typique d'interaction, ce qui est intuitivement le délai dans la transmission d'information d'un neurone à un autre. Une idée mieux adaptée, s'inspirant du chapitre 4, serait d'agréger plusieurs tests basés sur différentes valeurs de δ prises par exemple dans une grille, en tenant compte de la multiplicité des tests, qui rejetterait l'indépendance si au moins l'un des test individuel le fait. En particulier, le (ou les) délai δ pour le(s)quel la procédure agrégée rejette l'indépendance est alors disponible.

  Un point de départ pour l'analyse statistique d'une telle procédure est l'inégalité de concentration de type Bernstein obtenue dans le chapitre 3. Pour chaque δ, une borne suffisamment fine pour le quantile conditionnel correspondant en est déduite, menant à des conditions sur l'alternative garantissant un contrôle du risque de seconde espèce. Une étude par simulation devra également être faite afin de valider une telle procédure d'un point de vue pratique.

- Il serait également intéressant d'étudier d'autre types de fonctions φ. En particulier, un autre choix éventuel pour φ est de considérer la statistique introduite par Sansonnet et Tuleau-Malot dans [160], et de combiner l'approche par permutation à une procédure d'agrégation de tests basée sur leur choix de φ afin de voir si les nouvelles vitesses de séparation correspondent bien à celles qu'elles ont obtenu sous des hypothèses restrictives de modèles (à savoir, des processus à interactions poissonniennes).

- Concernant la méthode *Permutation Unitary Events* introduite dans le chapitre 2, il reste encore la justification théorique de la procédure de tests multiples. En effet, pour

l'instant, le contrôle du taux de faux positifs (FDR) est uniquement garanti pour les cas très particuliers de processus de Poisson homogènes, de fenêtres deux-à-deux disjointes et n'impliquant que les tests unilatéraux par valeurs supérieures (ou par valeurs inférieures), de sorte que les tests individuels soient indépendants. Nous avons étudié la propriété de dépendances positives (PRDS) sur l'ensemble des indices correspondant aux hypothèses nulles vraies, mais elle ne semble pas être vérifiée dans notre contexte. Ainsi, même si notre procédure semble être performante en pratique, afin d'obtenir une méthode complètement justifiée théoriquement, il faudrait soit appliquer d'autres procédures, telles que celles considérées par Blanchard et Roquain dans [23], comme en particulier, la procédure step-up de Benjamini et Yekutieli [17] remplaçant α par $\alpha \times (\sum_{j=1}^{m} 1/j)^{-1}$ dans celle de Benjamini et Hochberg.

- Finalement, un des inconvénients majeurs de la méthode introduite ici est qu'elle permet de détecter les synchronisations uniquement entre deux neurones. Par conséquent, une nouvelle perspective consiste à généraliser notre méthode à trois neurones, ou plus, afin de détecter différents groupes fonctionnels de neurones interagissant. En particulier, ceci nécessite la généralisation de l'approche par permutation à plus de deux coordonnées, et l'utilisation du nombre de coïncidences avec délai introduit par Chevallier et Laloë dans [35]. Ensuite, dans la lignée de leur article, afin d'éviter le choix délicat du motif à étudier, l'idée est d'appliquer une nouvelle étape de tests multiples, chaque test correspondant à un motif. Une fois de plus, la validité d'une telle procédure de test multiple reste à étudier.

Dans des contextes encore plus généraux, Gretton et ses co-auteurs ont construit des tests d'indépendance basés sur des noyaux reproduisants (voir par exemple [65, 66]). Leur statistique de test est basée sur une U-statistique et les valeurs critiques sont obtenues en pratique par permutation. Il serait intéressant d'étudier leur procédure de test d'un point de vue non-asymptotique, comme dans cette thèse, et de comparer les vitesses de séparation obtenues aux nôtres. Cependant, ceci nécessiterait la généralisation des inégalités de concentration pour les sommes permutées aléatoirement aux U-statistiques permutées mentionnée ci-dessus.

Appendix A

Mathematical Tools

This appendix succinctly presents some elementary results used in this thesis.

Contents

A.1 Some convergence results 225
A.1.1 Convergence of random variables 225
A.1.2 Convergence of distributions 226
A.2 The point process setting and the Skorohod topology 229
A.2.1 Simulation of point processes by Ogata's Thinning 229
A.2.2 The Skorohod topology and the point process framework 230
A.3 Some basics on U-statistics 231
A.3.1 Definition 231
A.3.2 Convergence results for U-statistics 233
A.4 Wavelets and Besov spaces 234
A.4.1 Multiresolution analysis and wavelets 234
A.4.2 Two-dimensional standard and weak Besov bodies 236

A.1 Some convergence results

This section mainly refers to the two great reviews on convergence and asymptotic statistics written by Dudley [46] and van der Vaart [172]. In all this section, fix a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, a separable metric space $(\mathcal{Y}, d_{\mathcal{Y}})$, endowed with the σ-algebra of Borel sets $\mathcal{B}_{\mathcal{Y}}$. Recall that a random variable Y on $\mathcal{Y}$ is a measurable function $Y : \Omega \to \mathcal{Y}$, that is such that for all Borel set B in $\mathcal{B}_{\mathcal{Y}}$, its inverse image $Y^{-1}(B)$ belongs to $\mathcal{A}$.

A.1.1 Convergence of random variables

Definition A.1.1. Let $Y, Y_1, Y_2, \dots$ be some random variables on $\mathcal{Y}$.

- The sequence $(Y_n)_n$ is said to converge *almost surely* (a.s.) to a random variable Y, denoted $Y_n \xrightarrow[n\to+\infty]{a.s.} Y$, if $\mathbb{P}\left(\left\{\omega \in \Omega \ ; \ Y_n(\omega) \underset{n\to+\infty}{\longrightarrow} Y(\omega)\right\}\right) = 1$.
- The sequence $(Y_n)_n$ is said to converge *in probability* to a random variable Y, denoted $Y_n \xrightarrow[n\to+\infty]{\mathbb{P}} Y$, if for all $\varepsilon > 0$, $\mathbb{P}(d_{\mathcal{Y}}(Y_n, Y) > \varepsilon) \underset{n\to+\infty}{\longrightarrow} 0$.

It is well known that a.s. convergence always implies convergence in probability, yet the converse does not always hold. Nevertheless, a kind of converse exists in terms of subsequences, stated in the following theorem, extracted from Dudley's [46] book.

Theorem A.1.1 (Dudley, Theorem 9.2.1)**.** *For any random variables $Y, Y_1, Y_2, \ldots$ from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ into a separable metric space $(\mathcal{Y}, d_{\mathcal{Y}})$, the sequence $(Y_n)_n$ converges to Y in probability if and only if, for every subsequence $\left(Y_{n(k)}\right)_k$, there exists a subsubsequence $\left(Y_{n(k(l))}\right)_l$ which converges almost surely to Y.*

A.1.2 Convergence of distributions

These two previous notions of convergence mainly concern the random variables. The third one introduced below rather involves their distribution. Recall that the distribution Q of a random variable Y is the probability measure on $(\mathcal{Y}, \mathcal{B}_{\mathcal{Y}})$ defined for all Borel sets B in $\mathcal{B}_{\mathcal{Y}}$ by $Q(B) = \mathbb{P}(Y \in B)$.

Definition A.1.2. A sequence $(Q'_n)_n$ of measures is said to converge *weakly* to a measure Q' (denoted by $Q'_n \underset{n\to+\infty}{\Longrightarrow} Q'$) if for all continuous and bounded function $g : \mathcal{Y} \to \mathbb{R}$,

$$\int_{\mathcal{Y}} g dQ'_n \underset{n\to+\infty}{\longrightarrow} \int_{\mathcal{Y}} g dQ'.$$

A sequence of random variables $(Y_n)_n$ is thus said to converge in distribution to a random variable Y if the sequence of their respective distributions $(Q_n)_n$ weakly converges to the distribution Q of Y, and in that case, we denote $Y_n \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} Y$.

Several links exist between the convergence of the random variables, and the convergence of their distribution. For instance, convergence in distribution is always implied by convergence in probability, and thus by almost sure convergence. Conversely, given a sequence of distributions which converge weakly, the Skorohod representation Theorem allows to construct random variables with such distributions which converge almost surely, as stated in the following Theorem extracted from Dudley's book [46, Theorem 11.7.2]. This theorem is very useful in bootstrap studies.

Theorem A.1.2 (Skorohod's representation theorem)**.** *Let $\mathcal{Y}$ be a separable metric space, and $Q, Q_1, Q_2, \ldots$ be some probability measures on $\mathcal{Y}$ such that $Q_n \underset{n\to+\infty}{\Longrightarrow} Q$.*
Then, there exists a probability space, and random variables $Y, Y_1, Y_2, \ldots$ on it with values in $\mathcal{Y}$ and respective distributions $Q, Q_1, Q_2, \ldots$, such that $Y_n \underset{n\to+\infty}{\xrightarrow{a.s.}} Y$.

Slutsky's lemma looks at the sequence of the joint distribution of two random variables and states the following.

Proposition A.1.1 (Slutsky's lemma)**.** *For any random variables $Y, Y_1, Y_2, \ldots$ and $Z_1, Z_2, \ldots$ from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ into a separable metric space $(\mathcal{Y}, d_{\mathcal{Y}})$, and any constant c in $\mathcal{Y}$, both $\left(Y_n \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} Y\right)$ and $\left(Z_n \underset{n\to+\infty}{\xrightarrow{\mathbb{P}}} c\right)$ imply that $\left((Y_n, Z_n) \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} (Y, c)\right)$.*

For convergence theorems in $\mathbb{R}^d$ with $d \geq 2$, one can reduce the study to the one-dimensional case thanks to the Cramér-Wold device, proved in [21, Theorem 29.4, p. 838] recalled below.

Theorem A.1.3 (The Cramér-Wold device)**.** *Let $Y = (Y^1, \ldots, Y^d)$, and $Y_n = (Y_n^1, \ldots, Y_n^d)$, $n \geq 1$ be some random vectors in $\mathbb{R}^d$. Then, $Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} Y$ if and only if*

$$\forall t = (t_1, \ldots, t_d) \in \mathbb{R}^d, \quad \sum_{k=1}^{d} t_k Y_n^k \xrightarrow[n\to+\infty]{\mathcal{L}} \sum_{k=1}^{d} t_k Y^k.$$

Weak convergence and cumulative distribution functions

In the real case $\mathcal{Y} = \mathbb{R}$, the weak convergence can be characterized in terms of the cumulative distribution functions (c.d.f.). Recall that the c.d.f. of a random variable Y with distribution Q is the function $F : t \in \mathbb{R} \mapsto \mathbb{P}(Y \leq t) = \int_{\mathbb{R}} \mathbb{1}_{(-\infty,t]} dQ$.
For all n, denote F_n (respectively F) the c.d.f. of Y_n (respectively Y). Then,

$$Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} Y \text{ if and only if } F_n(t) \xrightarrow[n\to+\infty]{} F(t) \text{ for all } t \text{ at which } F \text{ is continuous.}$$

Moreover, van der Vaart proves in [172] that the convergence is uniform as soon as the limit is continuous as stated in the lemma below.

Lemma A.1.1 (Van der Vaart, Lemma 2.11)**.** *Let $Y, Y_1, Y_2, \ldots$ be real-valued random variables, and $F, F_1, F_2, \ldots$ denote their corresponding c.d.f.'s. If $Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} Y$, and F is continuous, then* $\sup_{t\in\mathbb{R}} |F_n(t) - F(t)| \xrightarrow[n\to+\infty]{} 0$.

Moreover, weak convergence and quantile functions are also closely related. Recall that the quantile function q of a random variable Y is the generalized inverse of it's c.d.f., defined for all $\eta \in (0,1)$ by $q(\eta) = \inf\{t \in \mathbb{R} \ ; \ F(t) \geq \eta\}$. Then, van der Vaart characterizes the weak convergence in [172] as follows.

Lemma A.1.2 (Van der Vaart, Lemma 21.2)**.** *Let $Y, Y_1, Y_2, \ldots$ be real-valued random variables, and $q, q_1, q_2, \ldots$ denote their corresponding quantile functions. Then, $Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} Y$ if and only if $q_n(\eta) \xrightarrow[n\to+\infty]{} q(\eta)$ for all η at which q is continuous.*

In particular, if the limit Y has a continuous c.d.f., then $q_n(\eta) \xrightarrow[n\to+\infty]{} q(\eta)$ for all η in $(0,1)$.

Metrics for the weak convergence

Bounded Lipschitz metric. The weak convergence is metrizable, for instance by the *bounded Lipschitz metric* defined in [46, p. 394]. More precisely, let

$$\|f\|_{BL} = \sup_{u\neq v} \frac{|f(u) - f(v)|}{d_{\mathcal{Y}}(u,v)} + \|f\|_{\infty},$$

and denote BL the set of bounded Lipschitz functions on $\mathcal{Y}$. Then, the bounded Lipschitz metric between two probability measures Q and Q' on a separable metric space $(\mathcal{Y}, d_{\mathcal{Y}})$ is defined by

$$d_{BL}(Q, Q') = \sup_{f\in BL,\ \|f\|_{BL}\leq 1} \left| \int_{\mathcal{Y}} f dQ - \int_{\mathcal{Y}} f dQ' \right|.$$

Dudley proves in [46, Proposition 11.3.2] that d_{BL} is indeed a metric, that is, it satisfies the following properties:

(i) $d_{BL}(Q, Q') = 0$ if and only if $Q = Q'$,
(ii) $d_{BL}(Q, Q') = d_{BL}(Q', Q)$,
(iii) for all measures Q, Q', Q'', $d_{BL}(Q, Q'') \leq d_{BL}(Q, Q') + d_{BL}(Q', Q'')$.

He also proves in [46, Theorem 11.3.3] that d_{BL} metrizes the weak convergence, that is

$$Q_n \underset{n\to+\infty}{\Longrightarrow} Q \quad \text{if and only if} \quad d_{BL}(Q_n, Q) \underset{n\to+\infty}{\longrightarrow} 0.$$

Wasserstein's metric. Another metric which characterizes weak convergence is the Wasserstein metric. Here, we only consider the Wasserstein metric of order 2. Assume $(\mathcal{Y}, \|\cdot\|)$ is a separable Banach space, and let $\Gamma_2 = \Gamma_2(\mathcal{Y})$ be the set of probability measures Q on $\mathcal{Y}$ with finite second-order moments, that is

$$\Gamma_2 = \left\{ Q \text{ probability measure on } \mathcal{Y} \text{ such that } \int_{\mathcal{Y}} \|y\|^2 \, dQ(y) < +\infty \right\}.$$

The $\mathbb{L}_2$-Wasserstein metric, also called Mallows' metric, is defined between two probability measures Q and Q' in Γ_2 by

$$d_2^2(Q, Q') = \inf \left\{ \mathbb{E}\left[\|Y - Y'\|^2\right], \ (Y, Y') \text{ with marginals } Q \text{ and } Q' \right\}. \tag{A.1.1}$$

Let us recall two important properties of the Wasserstein metric which are proved by Bickel and Freedman in [20, Lemma 8.1 and Lemma 8.3].

Proposition A.1.2 (Bickel and Freedman, 1981)**.**

- *The infimum in* (A.1.1) *is attained, and* d_2 *is a metric on* Γ_2.
- *Let* Q *and* $Q_1, Q_2, \ldots$ *belong to* Γ_2. *Then, the two following statements are equivalent:*

  a) $d_2(Q_n, Q) \underset{n\to+\infty}{\longrightarrow} 0$,

  b) $Q_n \underset{n\to+\infty}{\Longrightarrow} Q$ *and* $\displaystyle\int_{\mathcal{Y}} \|y\|^2 \, dQ_n(y) \underset{n\to+\infty}{\longrightarrow} \int_{\mathcal{Y}} \|y\|^2 \, dQ(y)$.

Convergence of empirical measures

Usually, the convergence of the empirical measures is fundamental to prove consistency results for bootstrap approaches. In particular, the following theorem, extracted from Varadarajan's article [174, Theorems 1 and 3] guarantees the weak convergence of the empirical measures (almost surely in the sample) as soon as the data set $\mathcal{Y}$ is separable.

Theorem A.1.4 (Varadarajan, 1958)**.** *Let* $\mathcal{Y}$ *be a separable space and* $Y_1, Y_2, \ldots$ *be some independent and identically distributed (i.i.d.) random variables with distribution* Q *in* $\mathcal{Y}$. *Then*

$$\frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i} \underset{n\to+\infty}{\Longrightarrow} Q, \quad \textit{a.s. in } (Y_i)_i.$$

Moreover, in the real case $\mathcal{Y} = \mathbb{R}$, the following inequality, known as the Dvoretzky-Kiefer-Wolfowitz inequality, proved by Massart in [122, Corollary 1] (with the optimal constant) provides the convergence rate of the empirical distribution function to the underlying c.d.f. in $\mathbb{L}_\infty$-norm.

Theorem A.1.5 (Dvoretzky-Kiefer-Wolfowitz inequality)**.** *Let $Y_1, Y_2, \dots$ be i.i.d. real-valued random variables with common c.d.f. denoted F. Denote $\hat{F}_n$ their empirical distribution function, that is defined for all t in $\mathbb{R}$ by $\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{Y_i \leq t}$. Then for all $n \geq 1$ and all $x > 0$,*

$$\mathbb{P}\left(\sqrt{n}\sup_{t\in\mathbb{R}}\left|\hat{F}_n(t) - F(t)\right| > x\right) \leq 2\exp\left(-2x^2\right).$$

A.2 The point process setting and the Skorohod topology

A.2.1 Simulation of point processes by Ogata's Thinning

As explained in the introduction of this thesis, one may define a point process thanks to its conditional intensity (see Section 0.6.2), which is a function depending on the history of the process. Here, we focus on point processes supported on a compact interval $[0, T]$, where $0 < T < +\infty$, whose intensity is almost surely finite and in $\mathbb{L}_1([0, T])$.
Given such conditional intensity λ, it is possible to construct a point process on $[0, T]$ with such intensity in a simple way thanks to Ogata's [129] thinning procedure, generalizing Lewis and Shedler's [119] method to simulate inhomogenous Poisson processes. The idea is the following:

- Let $\mathcal{N}$ be a homogenous Poisson process with rate equal to 1 on $(\mathbb{R}_+)^2$, that is a point process such that, for all $k \geq 1$ and for all measurable subset $A_1, \dots, A_k$ of $(\mathbb{R}_+)^2$, the numbers of points of $\mathcal{N}$ in each of them, namely $\mathcal{N}(A_1), \dots, \mathcal{N}(A_k)$ are independent Poisson random variables with respective parameters the Lebesgue measure of $A_1, \dots, A_k$.

- The dependency with respect to the history of the process implies that the appearance of a new point on the process may reshape the form of the intensity which then has to be updated. The idea is thus to draw the graph of the conditional intensity starting from zero (from the left to the right). This is possible since the intensity is known until the first point of the process. Then, consider the first time the graph of the intensity passes on top of a point of $\mathcal{N}$, that is there exists a point in $\mathcal{N}$, say (T_1, U_1) such that $\lambda(T_1) \geq U_1$. Then, include T_1 to the point process X, and update consequently the intensity (since the history of the point process changes by adding T_1).

- Repeat the previous step. Draw the graph of the (updated) conditional intensity (given $\{T_1\}$) starting from T_1 and consider the first point, say (T_2, U_2), in $\mathcal{N}$ such that $\lambda(T_2) \geq U_2$. Include T_2 to the point process and update the conditional intensity (depending know on T_1 and T_2).

- Iterate this procedure until the end of the support interval $[0, T]$.

The fact that this procedure does indeed construct a point process with desired prescribed intensity can be found in [29] under reasonable assumptions, and a complete proof is detailed in [34].

A great advantage of this method is that one may derive properties from simple homogeneous Poisson processes to much more complicated processes as soon as they can be constructed by thinning.

A.2.2 The Skorohod topology and the point process framework

The definitions together with the properties concerning the Skorohod topology are introduced and detailed in [22, Chapter 3]. Let

$$\mathcal{D} = \{f : [0,1] \to \mathbb{R},\ f \text{ right-continuous and limited on the left}\},$$

that is the set of functions f satisfying:

- for all t_0 in $[0,1)$, $\lim_{\{t\to t_0, t>t_0\}} f(t)$ exists and is equal to $f(t_0)$,
- for all t_0 in $(0,1]$, $\lim_{\{t\to t_0, t<t_0\}} f(t)$ exists.

Definition A.2.1. Let Λ denote the set of functions $\lambda : [0,1] \to [0,1]$ that are continuous and (strictly) increasing, with $\lambda(0) = 0$ and $\lambda(1) = 1$. For all f, g in $\mathcal{D}$, define

$$d_{\mathcal{D}}(f,g) = \inf\left\{\varepsilon > 0\ ;\ \exists \lambda \in \Lambda \text{ satisfying both } \left\{ \begin{array}{l} \sup_{t\in[0,1]} |\lambda(t) - t| \le \varepsilon, \\ \sup_{t\in[0,1]} |f(\lambda(t)) - g(t)| \le \varepsilon, \end{array} \right\}\right.$$

Comments. Notice that when the $\mathbb{L}_\infty$-metric between two functions f and g, that is $\|f-g\|_\infty = \sup_{t\in[0,1]} |f(t) - g(t)|$, is small, it means that at fixed abscissa, the ordinates are very close. Yet here, in $d_{\mathcal{D}}(f,g)$, a little perturbation of the abscissa is authorized thanks to the deformation λ of the time scale. This embodies the idea that, for instance, time cannot be measured with perfect accuracy and thus, the resulting errors need to be taken into account.

Proposition A.2.1 (Billingsley, 2009)**.**

- *d is a metric on $\mathcal{D}$. It defines the Skorohod topology.*
- *$(\mathcal{D}, d_{\mathcal{D}})$ is a separable space.*

The proof of this proposition can be found in [22, p. 111-112].

The main reason to use the Skorohod topology in the point process framework is that it provides a distance between point processes with the great advantage of furnishing separability. More precisely, let $\mathcal{X}$ denote the set of all countable subsets of $[0,1]$. For all element x in $\mathcal{X}$, define its associated counting measure $dN_x = \sum_{u\in x} \delta_u$ as the sum of all delta measures concentrated at each point of x.
First, we embed $\mathcal{X}$ in the set $\mathcal{D}$ through the identification defined for all x in $\mathcal{X}$ by

$$N : x \in \mathcal{X} \mapsto \left(N_x : t \mapsto \int_0^1 \mathbb{1}_{u\le t}\ dN_x(u)\right).$$

In particular, $N_x(t)$ is equal to the number of points of x that occur in $[0,t]$. For all x, N_x belongs to $\mathcal{D}$. Moreover, it is non-decreasing, piecewise constant with jumps equal to one in each abscissa corresponding to a point of x. In fact, N_x is the counting process associated with x (see, e.g., [28]) in the point process theory.
Thanks to the identification N above, $\mathcal{X}$ can then be endowed with the topology induced by the metric $d_{\mathcal{X}}$ defined between to elements x and x' in $\mathcal{X}$ by

$$d_{\mathcal{X}}(x,x') = d_{\mathcal{D}}(N(x), N(x')).$$

Notice that the little deformation in $d_{\mathcal{D}}$ is what makes this metric very attractive for the point process framework. Indeed, consider a very simple example, where $x = \{1/2\}$ and $x_\eta = \{1/2 + \eta\}$, with $0 < \eta < 1/2$. Intuitively, if η tends to zero, we want to say that the distance between x and x_η tends to zero. If we had rather considered the $\mathbb{L}_\infty$-metric between the counting processes, then the distance between x and x_η would always be equal to one what ever $\eta > 0$ (since $N_x(1/2 + \eta/2) - N_{x_\eta}(1/2 + \eta/2) = 1$). Whereas, thanks to the deformation λ in the definition of $d_{\mathcal{D}}$, one can make the distance between x and x' smaller than η. Indeed, let

$$\lambda_\eta : t \in [0,1] \mapsto \begin{cases} \dfrac{1/2}{1/2+\eta} t & \text{if } 0 \le t \le 1/2 + \eta, \\ \\ \dfrac{1/2}{1/2-\eta}(t-1) + 1 & \text{if } 1/2 + \eta \le t \le 1. \end{cases}$$

This deformation consists in transforming the interval $[0, 1/2+\eta]$ into $[0, 1/2]$, and the interval $[1/2 + \eta]$ into $[1/2, 1]$. Then $\lambda_\eta \in \Lambda$, and

$$\begin{cases} \sup\limits_{t\in[0,1]} |\lambda_\eta(t) - t| = \eta, \\ \sup\limits_{t\in[0,1]} |N_x(\lambda_\eta(t)) - N_{x'}(t)| = 0, \end{cases}$$

hence $d_{\mathcal{X}}(x, x') \le \eta$.

Thanks to product metrics, we are able to extend the metric $d_{\mathcal{X}}$ to the product space $\mathcal{X}^2$ by letting for all couples $x = (x^1, x^2)$ and $x' = (x'^1, x'^2)$ in $\mathcal{X}^2$,

$$d_{\mathcal{X}^2}(x, x') = \sup_{j=1,2} \left\{ d_{\mathcal{X}}(x^j, x'^j) \right\},$$

and then extend it again on the product space $\mathcal{X}^2 \times \mathcal{X}^2$ by letting for all couples $x = (x^1, x^2)$, $y = (y^1, y^2)$, $x' = (x'^1, x'^2)$ and $y' = (y'^1, y'^2)$ in $\mathcal{X}^2$,

$$d\big((x, y), (x', y')\big) = \sup \left\{ \sup_{j=1,2} \left\{ d_{\mathcal{X}}(x^j, x'^j) \right\}, \sup_{j=1,2} \left\{ d_{\mathcal{X}}(y^j, y'^j) \right\} \right\}.$$

Finally, according to [46, Problem 3 p. 42], since $(\mathcal{D}, d_{\mathcal{D}})$ is a separable metric space, so are $(\mathcal{X}, d_{\mathcal{X}})$, $\big(\mathcal{X}^2, d_{\mathcal{X}^2}\big)$ and $(\mathcal{X}^2 \times \mathcal{X}^2, d)$.

A.3 Some basics on U-statistics

A.3.1 Definition

U-statistics are great tools for distribution free unbiased estimation. The theory of U-statistics, first developed by Hoeffding in [81], has been vastly investigated thereafter (see, for instance, the books of van der Vaart [172] or de la Peña and Giné [40], or the more recent one of Serfling [164]).

Definition A.3.1. Given a sample $\mathbb{Y}_n = (Y_1, \ldots, Y_n)$ of n i.i.d. random variables with distribution Q on $\mathcal{Y}$, the U-statistic of order r and kernel $h : \mathcal{Y}^r \to \mathbb{R}$ (that is a measurable function) is defined by

$$U_{n,h}(\mathbb{Y}_n) = \frac{(n-r)!}{n!} \sum_{(i_1,\ldots,i_r)\in I_n^r} h(Y_{i_1}, \ldots, Y_{i_r}),$$

where $I_n^r = \{(i_1, \dots, i_r) \in \{1, \dots, n\}^r \ ; \ \forall k \neq l, \ i_k \neq i_l\}$.

First notice that U-statistics of order one are classical sums of i.i.d. random variables. Moreover, noticing that $U_{n,h_{sym}}(\mathbb{Y}_n) = U_{n,h}(\mathbb{Y}_n)$ where

$$h_{sym} : (y_1, \dots, y_r) \in \mathcal{Y}^r \mapsto \frac{1}{r!} \sum_{\pi \in \mathfrak{S}_r} h(y_{\pi(1)}, \dots, y_{\pi(r)}), \tag{A.3.1}$$

one can assume without loss of generality that the kernel h is symmetric in his r entries, that is, for all $y_1, \dots, y_r$ in $\mathcal{Y}$, and for all permutation π of $\{1, \dots, r\}$, $h(y_{\pi(1)}, \dots, y_{\pi(r)}) = h(y_1, \dots, y_r)$. In this case, one can rewrite the U-statistic as

$$U_{n,h}(\mathbb{Y}_n) = \frac{1}{\binom{n}{r}} \sum_{1 \leq i_1 < \dots < i_r \leq n} h(Y_{i_1}, \dots, Y_{i_r}),$$

A main advantage of U-statistics is that they provide great distribution free unbiased estimators. In particular, if one wants to estimate a parameter $\theta(Q)$ that can be written as $\int_{\mathcal{Y}^r} g(y_1, \dots, y_r) dQ(y_1) \dots dQ(y_r)$ (with g an integrable function with respect to the product measure dQ^r), then by introducing the symmetrization $h = g_{sym}$ of g as in (A.3.1), one directly obtains that $U_{n,h}(\mathbb{Y}_n)$ is an unbiased estimator of $\theta(Q)$, since

$$\begin{aligned} \mathbb{E}\left[U_{n,h}(\mathbb{Y}_n)\right] &= \int_{\mathcal{Y}^r} h(y_1, \dots, y_r) \ dQ(y_1) \dots dQ(y_r) \\ &= \frac{1}{r!} \sum_{\pi \in \mathfrak{S}_r} \int_{\mathcal{Y}^r} g(y_1, \dots, y_r) \ dQ(y_1) \dots dQ(y_r) = \theta(Q). \end{aligned}$$

Let us introduce some examples.

- *Estimating the moments.* For $\mathcal{Y} = \mathbb{R}$, $r = 1$ and kernels $h : y \in \mathbb{R} \mapsto y^m$ for $m \geq 1$, we find the empirical mean ($m = 1$), and the empirical moments of order $m \geq 2$, that are

$$\frac{1}{n} \sum_{i=1}^n Y_i \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^n Y_i^m \text{ for } m \geq 2.$$

- *Estimating of the variance.* Let $\mathcal{Y} = \mathbb{R}$, and Y be a random variable with distribution Q on $\mathbb{R}$. Then, given a sample $\mathbb{Y}_n$ of n i.i.d. copies of Y, the U-statistic $U_{n,h}(\mathbb{Y}_n)$ or order $r = 2$ and kernel $h : (y_1, y_2) \in \mathbb{R}^2 \mapsto (y_1 - y_2)^2/2$ is a good estimator of the variance σ^2, since

$$\sigma^2 = \int_{\mathbb{R}^2} (y_1^2 - y_1 y_2) \ dQ(y_1) dQ(y_2).$$

- *Testing independence.* Let $\mathcal{Y} = \mathcal{X}^2$ be a product space, and Q be a distribution on $\mathcal{X}^2$ with marginals Q^1 and Q^2. When testing independence $Q = Q^1 \otimes Q^2$, one may be interested in testing the nullity of the parameter

$$\begin{aligned} \theta_\varphi(Q) = \int_{\mathcal{X}^2} \varphi(x^1, x^2) \left[dQ(x^1, x^2) - dQ^1(x^1) dQ^2(x^2)\right] \\ = \int_{\mathcal{X}^2 \times \mathcal{X}^2} \left[\varphi(x^1, x^2) - \varphi(x^1, x'^2)\right] dQ(x^1, x^2) dQ(x'^1, x'^2), \end{aligned}$$

for well-chosen measurable functions φ. The symmetrization step leads to the kernel defined for all $x = (x^1, x^2)$ and $x' = (x'^1, x'^2)$ in $\mathcal{X}^2 \times \mathcal{X}^2$ by

$$h_\varphi(x, x') = \frac{1}{2}\left[\varphi(x^1, x^2) - \varphi(x^1, x'^2) - \varphi(x'^1, x^2) - \varphi(x'^1, x'^2)\right],$$

and given a sample $\mathbb{Y}_n$ of n i.i.d. random variables with distribution Q on $\mathcal{X}^2$, an unbiased estimator of $\theta_\varphi(Q)$ is given by $U_{n,h_\varphi}(\mathbb{Y}_n)$.

A.3.2 Convergence results for U-statistics

In this section, we present some convergence theorems for U-statistics. For a better readability, let us denote $\mathbb{E}[h] = \int_{\mathcal{Y}^r} h(y_1, \ldots, y_r)\, dQ(y_1) \ldots dQ(y_r)$, and for all $l \leq r$,

$$\mathbb{E}[h|Y_1, \ldots, Y_l] = \int_{\mathcal{Y}^{r-l}} h(Y_1, \ldots, Y_l, y_{l+1}, \ldots, y_r)\, dQ(y_{l+1}) \ldots dQ(y_r).$$

Definition A.3.2. A U-statistic of order r and kernel $h : \mathcal{Y}^r \to \mathbb{R}$ is said to be degenerate if $\mathbb{E}[h|Y_1, \ldots, Y_{r-1}]$ is almost surely constant, that is

$$\mathrm{Var}(\mathbb{E}[h|Y_1, \ldots, Y_{r-1}]) = 0.$$

It is said to be non-degenerate otherwise.

At the basis of the convergence theorems for U-statistics presented here, is Hoeffding's decomposition [81]. It consists in writing the centered U-statistic as a sum of degenerate U-statistics with nice martingale properties (see for instance [85]) which are useful for proving convergence theorems.
More precisely, for all $1 \leq k \leq r$, there exists a symmetric measurable function $h_k : \mathcal{Y}^k \to \mathbb{R}$ satisfying $\mathbb{E}[h_k(Y_1, \ldots, Y_k)|Y_1, \ldots, Y_{k-1}] = 0$, such that

$$U_{n,h}(\mathbb{Y}_n) - \mathbb{E}[h] = \sum_{k=1}^{r} \binom{r}{k} U_{n,h_k}(\mathbb{Y}_n).$$

Moreover, for all k in $\{1, \ldots, r\}$, the function h_k is given by

$$h_k(Y_1, \ldots, Y_k) = (-1)^k \mathbb{E}[h] + \left(\sum_{l=1}^{k} (-1)^{k-l} \sum_{1 \leq i_1 < \cdots < i_l \leq k} \mathbb{E}[h|Y_{i_1}, \ldots, Y_{i_l}] \right).$$

For instance, in the particular case $r = 2$, Hoeffding's decomposition is exactly

$$U_{n,h}(\mathbb{Y}_n) - \mathbb{E}[h] = U_{n,h_2}(\mathbb{Y}_n) + 2U_{n,h_1}(\mathbb{Y}_n),$$

where $h_1(Y_1) = \mathbb{E}[h|Y_1] - \mathbb{E}[h]$ and $h_2(Y_1, Y_2) = h(Y_1, Y_2) - \mathbb{E}[h|Y_1] - \mathbb{E}[h|Y_2] + \mathbb{E}[h]$.

Law of large numbers

Theorem A.3.1 (Law of large numbers, Hoeffding 1961)**.**
Let $Y_1, Y_2, \ldots$ be i.i.d. random variables with distribution Q on $\mathcal{Y}$, $r \geq 1$, and $h : \mathcal{Y}^r \to \mathbb{R}$ be a symmetric measurable function. If $\mathbb{E}[|h(Y_1, \ldots, Y_r)|] < +\infty$, then

$$U_{n,h}(Y_1, \ldots, Y_n) \xrightarrow[n \to +\infty]{a.s.} \mathbb{E}[h].$$

The convergence is shown to hold in $\mathbb{L}_1$ under the same assumptions in [40, Theorem 4.1.4]. If $r = 1$, one recovers the classical law of large numbers for sums of i.i.d. random variables. Yet, when $r \geq 2$, the U-statistic is not an i.i.d. sum anymore. The proof is based on more complex arguments coming from the martingale theory (as did Hoeffding for instance in [85]).

Central limit theorem

As for i.i.d. sums, the following central limit theorem holds for non-degenerate U-statistics. The following theorem is a particular case of Hoeffding's theorem [81, Theorem 7.1] (since only one U-statistic is considered here).

Theorem A.3.2 (Central limit theorem, Hoeffding 1948)**.**
Let $Y_1, Y_2, \ldots$ be i.i.d. random variables with distribution Q on $\mathcal{Y}$, $r \geq 1$, and $h : \mathcal{Y}^r \to \mathbb{R}$ be a symmetric measurable function. If the U-statistic is non-degenerate and $\mathbb{E}\left[h^2\right] < +\infty$, then

$$\sqrt{n}(U_{n,h}(Y_1, \ldots, Y_n) - \mathbb{E}\left[h\right]) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \mathrm{Var}(\mathbb{E}\left[h|Y_r\right])\right).$$

This theorem is also proved in the particular case $r = 2$ in Chapter 1 as a by-product of Proposition 1.3.5.

A.4 Wavelets and Besov spaces

Part of the following material is borrowed from the thesis of Sansonnet [159], and the books of Meyer [126] or Daubechies [39] (both of them containing much more than what is presented here).
The multiresolution analysis leads to orthonormal bases whose elements are localized in both time and frequency domains. These bases are obtained from dyadic dilations and translations of a father wavelet ϕ and a mother wavelet ψ. Hence, let us denote for any wavelet function Φ in $\mathbb{L}_2(R)$, and for all j, k in $\mathbb{Z}$, the dilated at scale j and translated wavelet by

$$\Phi_{j,k} : x \in \mathbb{R} \mapsto 2^{j/2}\Phi(2^j x - k).$$

A.4.1 Multiresolution analysis and wavelets

One dimensional wavelet bases

The definition of wavelet bases is closely related to the notion of multiresolution analysis.

Definition A.4.1. A multiresolution analysis of $\mathbb{L}_2(\mathbb{R})$ is an increasing sequence $\{V_j\}_{j\in\mathbb{Z}}$, of linear subspaces of $\mathbb{L}_2(\mathbb{R})$ such that

- $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$, and $\bigcup_{j\in\mathbb{Z}} V_j$ is dense in $\mathbb{L}_2(\mathbb{R})$,
- for all f in $\mathbb{L}_2(\mathbb{R})$ and all j in $\mathbb{Z}$, $f(\cdot) \in V_j \Leftrightarrow f(2\times\cdot) \in V_{j+1}$,
- for all f in $\mathbb{L}_2(\mathbb{R})$ and all k in $\mathbb{Z}$, $f(\cdot) \in V_0 \Leftrightarrow f(\cdot - k) \in V_0$,
- there exists a function ϕ in V_0, named father wavelet or scaling function, such that $\{\phi(\cdot - k), k \in \mathbb{Z}\}$ is an orthonormal basis of V_0.

Hence, as a consequence, for each resolution level (or scale) j, the space V_j is generated by the orthonormal basis $\{\phi_{j,k}, k \in \mathbb{Z}\}$.
For all j, consider the orthogonal complement W_j of V_j in V_{j+1}, that is the subset of V_{j+1} such that $V_{j+1} = V_j \oplus W_j$. In particular, by definition, one obtains that

$$\mathbb{L}_2(\mathbb{R}) = V_0 \oplus \bigoplus_{j\geq 0} W_j. \tag{A.4.1}$$

A basic tenet of multiresolution analysis is that one can build a function ψ, called mother wavelet, such that for all j in $\mathbb{Z}$, $\{\psi_{j,k}, k \in \mathbb{Z}\}$ is an orthonormal basis of W_j. Hence, (A.4.1) implies that $\{\phi_{0,k}, \psi_{j,k}, j \geq 0, k \in \mathbb{Z}\}$ is an orthonormal basis of $\mathbb{L}_2(\mathbb{R})$, that is for all g in $\mathbb{L}_2(\mathbb{R})$, for all x in $\mathbb{R}$,

$$g(x) = \sum_{k\in\mathbb{Z}} \alpha_{0,k}(g)\phi_{0,k}(x) + \sum_{j\geq 0}\sum_{k\in\mathbb{Z}} \beta_{j,k}(g)\psi_{j,k}(x),$$

where the wavelet coefficients are defined by

$$\alpha_{0,k}(g) = \int_{\mathbb{R}} g(x)\phi_{0,k}(x)dx \quad \text{and} \quad \beta_{j,k}(g) = \int_{\mathbb{R}} g(x)\psi_{j,k}(x)dx.$$

One-dimensional Haar basis. As an example, one could consider the one-dimensional Haar basis defined by

$$\phi = \mathbb{1}_{[0,1]} \quad \text{and} \quad \psi = \mathbb{1}_{[0,1.2)} - \mathbb{1}_{[1/2,1)}.$$

In particular, if for all j, V_j denotes the topological closure of the linear subspace generated by $\{\phi_{j,k}, k \in \mathbb{Z}\}$, that is $V_j = \overline{\text{Span}\,\langle \phi_{j,k}, k \in \mathbb{Z}\rangle}$, the collection of sets $\{V_j\}_{j\in\mathbb{Z}}$ is a multiresolution analysis of $\mathbb{L}_2(\mathbb{R})$. Notice that the set V_j consists of all functions that are piecewise constant on each interval $[k2^{-j}; (k+1)2^{-j})$, $k \in \mathbb{Z}$.

Moreover notice that by considering only translations k in $\mathcal{K}_j = \{0, 1, \ldots, 2^j - 1\}$, one obtains an orthonormal basis of $\mathbb{L}_2([0,1])$, that is precisely $\{\phi\} \cup \{\psi_{j,k}, j \geq 0, k \in \mathcal{K}_j\}$. This is a particularity of Haar's basis and especially the fact that both ϕ and ψ are supported by $[0,1]$. When considering other bases of $\mathbb{L}_2(\mathbb{R})$ in order to recover bases of $\mathbb{L}_2([0,1])$, edge wavelets need to be added.

Two-dimensional wavelet bases

There are several approaches leading to the construction of two-dimensional multiresolution analyses from a one-dimensional one. For all functions g and h in $\mathbb{L}_2(\mathbb{R})$, define their tensor product by

$$g \otimes h : (y, z) \in \mathbb{R}^2 \mapsto g(y)h(z).$$

A first construction is to consider the collection of wavelets obtained from all tensor products of the elements in the one-dimensional basis. This leads to an anisotropic basis since the two coordinates are dilated separately.

Another construction, more interesting for many applications, is rather based on the products of one-dimensional multiresolution analysis spaces. This leads to an isotropic multiresolution of $\mathbb{L}_2(\mathbb{R}^2)$ (that is defined exactly as in Definition A.4.1, replacing $\mathbb{R}$ by $\mathbb{R}^2$ and "$k \in \mathbb{Z}$" by $k = (k_1, k_2) \in \mathbb{Z}^2$), the term "isotropic" meaning that the dilation in all the directions is the same.

The idea is to consider

$$\mathbf{V}_0 = V_0 \otimes V_0 = \overline{\text{Span}\,\langle g \otimes h;\ g, h \in V_0\rangle}, \quad \text{and} \quad f(\cdot,\cdot) \in \mathbf{V_j} \Leftrightarrow f(2^j\cdot, 2^j\cdot) \in \mathbf{V}_0.$$

Then $\{\mathbf{V}_j, j \in \mathbb{Z}\}$ is a multiresolution analysis of $\mathbb{L}_2(\mathbb{R}^2)$. Moreover, one can construct the corresponding two-dimensional wavelet basis from the one-dimensional one: denote for all $k = (k_1, k_2) \in \mathbb{Z}^2$,

$$\varphi_{0,k} : (y^1, y^2) \in \mathbb{R}^2 \mapsto \varphi_0(y^1 - k_1, y^2 - k_2), \quad \text{where} \quad \varphi_0 = \phi \otimes \phi,$$

then $\{\varphi_{0,k}, k \in \mathbb{Z}^2\}$ is an orthonormal basis of $\mathbf{V}_0$.
As for the one-dimensional case, one can consider $\mathbf{W}_j$ the orthogonal complement of $\mathbf{V}_j$ in $\mathbf{V}_{j+1}$, and in particular,

$$\mathbf{V}_{j+1} = V_{j+1} \otimes V_{j+1} = (V_j \otimes V_j) \oplus \underbrace{[(V_j \otimes W_j) \oplus (W_j \otimes V_j) \oplus (W_j \otimes W_j)]}_{\mathbf{W}_j} = \mathbf{V}_j \oplus \mathbf{W}_j.$$

The idea is thus to introduce three wavelets, denoted here by

$$\varphi_1 = \phi \otimes \psi, \quad \varphi_2 = \psi \otimes \phi \quad \text{and} \quad \varphi_3 = \psi \otimes \psi.$$

Then, for all j in $\mathbb{Z}$, if one denotes for all $1 \leq i \leq 3$, and all $k = (k_1, k_2)$ in $\mathbb{Z}^2$,

$$\varphi_{i,j,k} : \quad (y^1, y^2) \in \mathbb{R}^2 \quad \mapsto \quad \varphi_i(2^j y^1 - k_1, 2^j y^2 - k_2),$$

the family $\{\varphi_{i,j,k}, 1 \leq i \leq 3, k \in \mathbb{Z}^2\}$ is an orthonormal basis of $\mathbf{W}_j$ and therefore,

$$\{\varphi_{0,k}, k \in \mathbb{Z}^2\} \cup \{\varphi_{i,j,k}, 1 \leq i \leq 3, j \geq 0, k \in \mathbb{Z}^2\}$$

is an orthonormal basis of $\mathbb{L}_2(\mathbb{R}^2)$. Notice that this basis is indeed isotropic.

Two-dimensional Haar basis. As an example, the two-dimensional Haar basis is defined following the previous steps based on the one-dimensional Haar father wavelet $\phi = \mathbb{1}_{[0,1]}$ and mother wavelet $\psi = \mathbb{1}_{[0,1.2)} - \mathbb{1}_{[1/2,1)}$.

Moreover notice that, once again, by considering only translations k in $\mathcal{K}_j := \{0, 1, \ldots, 2^j - 1\}^2$, one obtains an orthonormal basis of $\mathbb{L}_2([0,1]^2)$, that is precisely

$$\{\varphi_0\} \cup \{\varphi_{i,j,k}, 1 \leq i \leq 2, j \geq 0, k \in \mathcal{K}_j\}. \tag{A.4.2}$$

A.4.2 Two-dimensional standard and weak Besov bodies

Besov spaces. Standard Besov spaces $\mathcal{B}^{\delta}_{p,q}$ are very general tools for describing the smoothness properties of functions in $\mathbb{L}_2$. In particular, they include well-known regularity spaces such as Hölder spaces (for $p = q = \infty$), or Sobolev spaces (for $p = q = 2$). Usually, Besov spaces with smoothness parameter $\delta > 0$ are defined thanks to the moduli of continuity of the derivatives of order $\lfloor \delta \rfloor$. A great advantage of these sets is that they can be characterized in terms of wavelet coefficients. Here, we only focus on these sequential characterizations. In this thesis, we only focus on Besov bodies with parameter $p = 2$ and $q = \infty$, but more general definitions can be found in [126, 76]. Moreover, as we work in $\mathbb{L}_2([0,1]^2)$, we consider the two-dimensional Haar basis defined in (A.4.2). Let us introduce the following definition inspired by [9].

Definition A.4.2. A function $g = \beta_0 \varphi_0 + \sum_{j \geq 0} \sum_{i=1}^{3} \sum_{k \in \mathcal{K}_j} \beta_{i,j,k} \varphi_{i,j,k}$ in $\mathbb{L}_2([0,1]^2)$ belongs to the Besov body $\mathcal{B}^{\delta}_{2,\infty}(R)$ of smoothness parameter δ and radius R if

$$\beta_0^2 \leq R^2 \quad \text{and} \quad \sup_{J \geq 0} \left\{ \sum_{j \geq J} \sum_{i=1}^{3} \sum_{k \in \mathcal{K}_j} \beta_{i,j,k}^2 \right\} \leq R^2.$$

Notice that one recovers standard Besov bodies provided that the father and mother wavelets are smooth enough (at least with smoothness parameter larger than $\delta - 1$). In particular, to do so, one should consider more regular wavelets, such as Daubechies' ones for instance, instead of considering the Haar basis.

Weak Besov spaces. Let us now introduce a weaker version of Besov spaces, namely the *weak Besov spaces*.
As above, consider the two-dimensional Haar basis of $\mathbb{L}_2([0,1]^2)$. A function $g = \beta_0\varphi_0 + \sum_{j\geq 0}\sum_{i=1}^{3}\sum_{k\in\mathcal{K}_j}\beta_{i,j,k}\varphi_{i,j,k}$ in $\mathbb{L}_2([0,1]^2)$ belongs to the weak Besov body $\mathcal{W}_\gamma(R')$ of smoothness parameter γ and radius R' if

$$\sup_{t>0}\left\{ t^{\frac{2}{\gamma+1}}\left(\mathbb{1}_{|\beta_0|>t} + \sum_{j\geq J}\sum_{i=1}^{3}\sum_{k\in\mathcal{K}_j}\mathbb{1}_{|\beta^2_{i,j,k}|>t}\right)\right\} \leq {R'}^{\frac{2}{\gamma+1}}. \tag{A.4.3}$$

Whereas standard Besov spaces constitute ideal classes to measure smoothness properties, $\mathcal{W}_\gamma$ constitute great spaces to measure the sparsity of functions in $\mathbb{L}_2([0,1]^2)$.
Indeed, following [152], one can make a link with the weak ℓ_p spaces that are defined by

$$w\ell_\nu = \left\{ (\theta_n)_{n\in\mathbb{N}}\ ;\ \sup_{t>0} t^\nu \sum_{n\in\mathbb{N}}\mathbb{1}_{|\theta_n|>t} < +\infty \right\}.$$

To do so, notice that a function g in $\mathbb{L}_2$ belongs to $\mathcal{W}_\gamma$ if and only if the left-hand side in (A.4.3) is finite, that is, the sequence of wavelet coefficients of g belongs to $w\ell_{2/(\gamma+1)}$. In particular, this condition is equivalent to saying that the number of large entries is controlled by a power law. One may moreover notice that it is also equivalent to $\sup_n n^{(\gamma+1)/2}\,|\theta|_{(n)} < +\infty$, where

$$|\theta|_{(1)} \geq |\theta|_{(2)} \geq \cdots \geq |\theta|_{(n)} \geq \ldots,$$

which is exactly Abramovich et al.s' [1] notion of sparsity for infinite vectors in $\mathbb{R}^{\mathbb{N}}$.

Bibliography

[1] F. Abramovich, Y. Benjamini, D. L. Donoho, and I. M. Johnstone. Adapting to unknown sparsity by controlling the false discovery rate. Technical report, Department of Statistics - Stanford University, 2000.

[2] R. Adamczak, D. Chafaï, and P. Wolff. Circular law for random matrices with exchangeable entries. *arXiv preprint arXiv:1402.3660*, 2014.

[3] A. M. Aertsen, G. L. Gerstein, M. K. Habib, and G. Palm. Dynamics of neuronal firing correlation: modulation of "effective connectivity". *Journal of neurophysiology*, 61(5):900–917, 1989.

[4] J. Antoch and M. Hušková. Permutation tests in change point analysis. *Statistics & Probability Letters*, 53(1):37–46, 2001.

[5] M. A. Arcones and E. Giné. On the bootstrap of U and V statistics. *The Annals of Statistics*, 20(2):655–674, 1992.

[6] M. A. Arcones and E. Giné. Limit theorems for U-processes. *The Annals of Probability*, 21(3):1494–1542, 1993.

[7] S. Arlot, G. Blanchard, and E. Roquain. Some nonasymptotic results on resampling in high dimension, I: Confidence regions and II: Multiple tests. *The Annals of Statistics*, 38(1):51–82, 83–99, 2010.

[8] J. N. Arvesen. Jackknifing U-statistics. *The Annals of Mathematical Statistics*, 40(6):2076–2100, 1969.

[9] F. Autin, E. Le Pennec, and K. Tribouley. Thresholding methods to estimate copula density. *Journal of Multivariate Analysis*, 101(1):200–222, 2010.

[10] O. Avila-Akerberg and M. J. Chacron. Nonrenewal spike train statistics: causes and functional consequences on neural coding. *Experimental brain research*, 210(3-4):353–371, 2011.

[11] E. Bacry, K. Dayri, and J. F. Muzy. Non-parametric kernel estimation for symmetric hawkes processes. application to high frequency financial data. *The European Physical Journal B*, 85(5):1–12, 2012.

[12] E. Bacry, S. Delattre, M. Hoffmann, and J. F. Muzy. Scaling limits for Hawkes processes and application to financial statistics. *Stochastic Process. Appl.*, 123(7):2475–2499, 2013.

[13] Y. Baraud. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606, 2002.

[14] Y. Baraud, S. Huet, and B. Laurent. Adaptive tests of linear hypotheses by model selection. *Annals of statistics*, pages 225–251, 2003.

[15] H. B. Barlow. Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1:371–394, 1972.

[16] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995.

[17] Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188, 2001.

[18] G. Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962.

[19] S. Bernstein. The theory of probabilities, 1946.

[20] P. J. Bickel and D. A. Freedman. Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9(6):1196–1217, 1981.

[21] P Billingsley. *Probability and measure*. Series in Probability and Statistics. Wiley, 9th edition, 1995.

[22] P. Billingsley. *Convergence of probability measures*. Wiley-Interscience, 2009.

[23] G. Blanchard and E. Roquain. Two simple sufficient conditions for fdr control. *Electronic journal of Statistics*, 2:963–992, 2008.

[24] J. R. Blum, J. Kiefer, and M. Rosenblatt. Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics*, 32(2):485–498, 1961.

[25] E. Bolthausen. An estimate of the remainder in a combinatorial central limit theorem. *Probability Theory and Related Fields*, 66(3):379–386, 1984.

[26] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, 2013.

[27] S. Bouzebda. General tests of independence based on empirical processes indexed by functions. *Statistical Methodology*, 21:59–87, 2014.

[28] P. Brémaud. *Point processes and queues*. Springer-Verlag, New York-Berlin, 1981. Martingale dynamics, Springer Series in Statistics.

[29] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *The Annals of Probability*, 24(3):1563–1588, 1996.

[30] J. Bretagnolle. Lois limites du bootstrap de certaines fonctionnelles. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 19(3):281–296, 1983.

[31] B. M. Brown. Martingale central limit theorems. *The Annals of Mathematical Statistics*, 42(1):59–66, 1971.

[32] C. Butucea and K. Tribouley. Nonparametric homogeneity tests. *Journal of statistical planning and inference*, 136(3):597–639, 2006.

[33] S. Chatterjee. Stein's method for concentration inequalities. *Probability theory and related fields*, 138(1):305–321, 2007.

[34] J. Chevallier, M. J. Caceres, M. Doumic, and P. Reynaud-Bouret. Microscopic approach of a time elapsed neural model. *arXiv preprint arXiv:1506.02361*, 2015.

[35] J. Chevallier and T. Laloë. Detection of dependence patterns with delay. *arXiv preprint arXiv:1406.0476*, 2014.

[36] E. Chung and J. P. Romano. Exact and asymptotically robust permutation tests. *The Annals of Statistics*, 41(2):484–507, 2013.

[37] A. Cohen, I. Daubechies, and P. Vial. Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 1(1):54–81, 1993.

[38] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. I.* Probability and its Applications. Springer-Verlag, New York, 2003.

[39] I. Daubechies. *Ten lectures on wavelets*, volume 61. SIAM, 1992.

[40] V. H. De la Peña and E. Giné. *Decoupling. Probability and its Applications.* Springer-Verlag, New York, 1999.

[41] H. Dehling and T. Mikosch. Random quadratic forms and the bootstrap for U-statistics. *Journal of Multivariate Analysis*, 51(2):392–413, 1994.

[42] B. Delyon. Exponential inequalities for sums of weakly dependent variables. *Electronic Journal of Probability*, 14:752–779, 2009.

[43] B. Delyon and A. Juditsky. On minimax wavelet estimators. *Applied and Computational Harmonic Analysis*, 3(3):215–228, 1996.

[44] M. Denker, B. Wiebelt, D. Fliegner, M. Diesmann, and A. Morrison. Practically trivial parallel data processing in a neuroscience laboratory. In *Analysis of Parallel Spike Trains*, pages 413–436. Springer Series in Computational Neuroscience, 2010.

[45] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Density estimation by wavelet thresholding. *The Annals of Statistics*, pages 508–539, 1996.

[46] R. M. Dudley. *Real Analysis and Probability.* Cambridge University Press, 2002.

[47] S. Dudoit and M. J. van der Laan. *Multiple testing procedures with applications to genomics.* Springer Science & Business Media, 2007.

[48] C. Durot and Y. Rozenholc. An adaptive test for zero mean. *Mathematical Methods of Statistics*, 15(1):26–60, 2006.

[49] M. Dwass. On the asymptotic normality of some statistics used in non-parametric tests. *The Annals of Mathematical Statistics*, pages 334–339, 1955.

[50] B. Efron. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979.

[51] B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1993.

[52] M. S. Ermakov. Minimax detection of a signal in a gaussian white noise. *Theory of Probability & Its Applications*, 35(4):667–679, 1991.

[53] F. Farkhooi, M. F. Strube-Bloss, and M. P. Nawrot. Serial correlation in neural spike trains: Experimental evidence, stochastic modeling, and single neuron variability. *Physical Review E*, 79(2):021905, 2009.

[54] R. A. Fisher. *The design of experiments*. Edinburgh & London: Oliver & Boyd, 1935.

[55] M. Fromont and B. Laurent. Adaptive goodness-of-fit tests in a density model. *The Annals of Statistics*, 34(2):680–720, 2006.

[56] M. Fromont, B. Laurent, M. Lerasle, and P. Reynaud-Bouret. Kernels based tests with non-asymptotic bootstrap approaches for two-sample problems. *Journal of Machine Learning Research: Workshop and Conference Proceedings, COLT 2012*, 23:23.1–23.22, 2012.

[57] M. Fromont, B. Laurent, and P. Reynaud-Bouret. Adaptive tests of homogeneity for a poisson process. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 47(1):176–213, 2011.

[58] M. Fromont, B. Laurent, and P. Reynaud-Bouret. The two-sample problem for poisson processes: Adaptive tests with a nonasymptotic wild bootstrap approach. *The Annals of Statistics*, 41(3):1431–1461, 2013.

[59] M. Fromont, M. Lerasle, and P. Reynaud-Bouret. Family wise separation rates for multiple testing. Technical report, https://hal.archives-ouvertes.fr/hal-01107321, Jan 2015.

[60] G. Gayraud and K. Tribouley. A goodness-of-fit test for copula densities. *test*, 20(3):549–573, 2011.

[61] G. L. Gerstein and D. H. Perkel. Simultaneous recorded trains of action potentials: analysis and functional interpretation. *Science*, 164:828–830, 1969.

[62] E. Gine. Lectures on some aspects of the bootstrap. In *Lecture Notes in Mathematics. École d'été de Probabilités de Saint-Flour, XXVI-1996*, pages 37–151. Springer, 1997.

[63] P. I. Good. *Permutation, parametric, and bootstrap tests of hypotheses*. Springer Science & Business Media, 2006.

[64] F. Grammont and A. Riehle. Spike synchronization and firing rate in a population of motor cortical neurons in relation to movement direction and reaction time. *Biological Cybernetics*, 88(5):360–373, 2003.

[65] A. Gretton, K. Fukumizu, C. H. Teo, L. Song, B. Schölkopf, and A. J. Smola. A kernel statistical test of independence. In *Advances in Neural Information Processing Systems*, pages 585–592, 2007.

[66] A. Gretton and L. Györfi. Consistent nonparametric tests of independence. *Journal of Machine Learning Research*, 11:1391–1423, 2010.

[67] S. Grün. *Unitary joint-events in multiple-neuron spiking activity: detection, significance, and interpretation*. Thun: Verlag Harri Deutsch, 1996.

[68] S. Grün. Data-driven significance estimation for precise spike correlation. *Journal of Neurophysiology*, 101:1126–1140, 2009.

[69] S. Grün, M. Diesmann, and A. Aertsen. Unitary events in multiple single-neuron spiking activity: I. detection and significance. *Neural Computation*, 14(1):43–80, 2002.

[70] S. Grün, M. Diesmann, and A. Aertsen. Unitary events in multiple single-neuron spiking activity: Ii. nonstationary data. *Neural computation*, 14(1):81–119, 2002.

[71] S. Grün, M. Diesmann, and A. Aertsen. Unitary event analysis. In *Analysis of parallel spike trains*, pages 191–220. Springer Series in Computational Neuroscience, 2010.

[72] S. Grün, M. Diesmann, F. Grammont, A. Riehle, and A. Aertsen. Detecting unitary events without discretization of time. *Journal of neuroscience methods*, 94(1):67–79, 1999.

[73] G. Gusto and S. Schbath. FADO: a statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes' model. *Statistical Applications in Genetics and Molecular Biology*, 4, 2005.

[74] J. Hájek. Some extensions of the wald-wolfowitz-noether theorem. *The Annals of Mathematical Statistics*, 32(2):506–523, 1961.

[75] N. R. Hansen, P. Reynaud-Bouret, and V. Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli*, 21(1):83–143, 2015.

[76] W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov. *Wavelets, approximation, and statistical applications*, volume 129. Springer Science & Business Media, 2012.

[77] M. T. Harrison, A. Amarasingham, and R. E. Kass. Statistical identification of synchronous spiking. In *Spike Timing: Mechanisms and Function*, page 77. CRC Press, 2013.

[78] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[79] D. O. Hebb. *The organization of behavior: A neuropsychological theory*. Wiley, New York, 1949.

[80] S. T. Ho and L. H. Y. Chen. An L_p bound for the remainder in a combinatorial central limit theorem. *The Annals of Probability*, 6(2):231–249, 1978.

[81] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.*, 19(3):293–325, 1948.

[82] W. Hoeffding. A non-parametric test of independence. *The Annals of Mathematical Statistics*, 19(4):546–557, 1948.

[83] W. Hoeffding. A combinatorial central limit theorem. *The Annals of Mathematical Statistics*, 22(4):558–566, 1951.

[84] W. Hoeffding. The large-sample power of tests based on permutation of the observations. *The Annals of Mathematical Statistics.*, 23(2):169–192, 1952.

[85] W. Hoeffding. The strong law of large numbers for U-statistics. *Institute of Statistics, Mimeograph series*, 302, 1961.

[86] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[87] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.

[88] L. Horváth and M. Hušková. Testing for changes using permutations of U-statistics. *Journal of Statistical Planning and Inference*, 128(2):351–371, 2005.

[89] H. Hotelling and M. R. Pabst. Rank correlation and tests of significance involving no assumption of normality. *The Annals of Mathematical Statistics*, 7(1):29–43, 1936.

[90] M. Hušková and P. Janssen. Generalized bootstrap for studentized U-statistics: A rank statistic approach. *Statistics & probability letters*, 16(3):225–233, 1993.

[91] Yu. I. Ingster. On the minimax nonparametric detection of signals in white gaussian noise. *Problemy Peredachi Informatsii*, 18(2):61–73, 1982.

[92] Yu. I. Ingster. Asymptotically minimax testing of nonparametric hypotheses. *Prob. Theory Math. Stat. I*, pages 553–574, 1987.

[93] Yu. I. Ingster. Minimax testing of nonparametric hypotheses on a distribution density in the L_p metrics. *Theory of Probability & Its Applications*, 31(2):333–337, 1987.

[94] Yu. I. Ingster. Asymptotic minimax testing of independence hypothesis. *Journal of Soviet Mathematics*, 44(4):466–476, 1989.

[95] Yu. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III. *Math. Methods Statist*, 2(2):85–114, 1993.

[96] Yu. I. Ingster. Minimax testing of the hypothesis of independence for ellipsoids inl p. *Journal of Mathematical Sciences*, 81(1):2406–2420, 1996.

[97] Yu. I. Ingster. Adaptive chi-square tests. *Journal of Mathematical Sciences*, 99(2):1110–1119, 2000.

[98] Yu. I. Ingster and Yu. A Kutoyants. Nonparametric hypothesis testing for intensity of the poisson process. *Mathematical Methods of Statistics*, 16(3):217–245, 2007.

[99] Yuri I Ingster and Irina A Suslina. Minimax nonparametric hypothesis testing for ellipsoids and besov bodies. *ESAIM: Probability and Statistics*, 4:53–135, 2000.

[100] A. Janssen and T. Pauls. How do bootstrap and permutation tests work? *The Annals of statistics*, 31(3):768–806, 2003.

[101] K. Jogdeo. Asymptotic normality in nonparametric methods. *The Annals of Mathematical Statistics*, pages 905–922, 1968.

[102] R. E. Kass, U. T. Eden, and E. N. Brown. *Analysis of neural data*. Springer, 2014.

[103] R. E. Kass, R. C. Kelly, and W.-L. Loh. Assessment of synchrony in multiple neural spike trains using loglinear point process models. *The annals of applied statistics*, 5(2B):1262, 2011.

[104] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1–2):81–93, 1938.

[105] M. G. Kendall, S. FH Kendall, and B. B. Smith. The distribution of spearman's coefficient of rank correlation in a universe in which all rankings occur an equal number of times. *Biometrika*, pages 251–273, 1939.

[106] G. Kerkyacharian and D. Picard. Density estimation in besov spaces. *Statistics & Probability Letters*, 13(1):15–24, 1992.

[107] G. Kerkyacharian and D. Picard. Density estimation by kernel and wavelets methods: optimality of besov spaces. *Statistics & Probability Letters*, 18(4):327–336, 1993.

[108] G. Kerkyacharian and D. Picard. Thresholding algorithms, maxisets and well-concentrated bases. *Test*, 9(2):283–344, 2000.

[109] B. E. Kilavik, S. Roux, A. Ponce-Alvarez, J. Confais, S. Grün, and A. Riehle. Long-term modifications in motor cortical dynamics induced by intensive practice. *The Journal of Neuroscience*, 29(40):12653–12663, 2009.

[110] C. Kirch. *Resampling Methods for the Change Analysis of Dependent Data*. PhD thesis, Universität zu Köln, 2006.

[111] C. Kirch and J. Steinebach. Permutation principles for the change analysis of stochastic processes under strong invariance. *Journal of Computational and Applied Mathematics*, 186(1):64–88, 2006.

[112] L. Klebanov, A. Gordon, Y. Xiao, H. Land, and A. Yakovlev. A permutation test motivated by microarray data analysis. *Computational Statistics & Data Analysis*, 50(12):3619–3628, 2006.

[113] M. Krumin, I. Reutsky, and S. Shoham. Correlation-based analysis and generation of multiple spike trains using hawkes models with an exogenous input. *Frontiers in computational neuroscience*, 4, 2010.

[114] M. Ledoux. *The concentration of measure phenomenon*, volume 89. American Mathematical Soc., 2005.

[115] O. V. Lepski and C. F. Pouet. Hypothesis testing under composite functions alternative. In *Topics in Stochastic Analysis and Nonparametric Estimation*, pages 123–150. Springer, 2008.

[116] O. V. Lepski and V. G. Spokoiny. Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli*, 5(2):333–358, 1999.

[117] O. V. Lepski and A. B. Tsybakov. Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probability Theory and Related Fields*, 117(1):17–48, 2000.

[118] A. Leucht and M. H. Neumann. Consistency of general bootstrap methods for degenerate U-type and V-type statistics. *Journal of Multivariate Analysis*, 100(8):1622–1633, 2009.

[119] P. A. Lewis and G. S. Shedler. Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, 1979.

[120] S. Louis, C. Borgelt, and S. Grün. Generation and selection of surrogate methods for correlation analysis. In *Analysis of Parallel Spike Trains*, pages 359–382. Springer, 2010.

[121] S. Louis, G. L. Gerstein, S. Grün, and M. Diesmann. Surrogate spike train generation through dithering in operational time. *Frontiers in computational neuroscience*, 4, 2010.

[122] P. Massart. The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *The Annals of Probability*, pages 1269–1283, 1990.

[123] P. Massart and J. Picard. *Concentration inequalities and model selection*, volume 1896. Springer, 2007.

[124] C. McDiarmid. Concentration for independent permutations. *Combinatorics, Probability & Computing*, 11(02):163–178, 2002.

[125] S. G. Meintanis and G. Iliopoulos. Fourier methods for testing multivariate independence. *Computational Statistics & Data Analysis*, 52:1884–1895, 2008.

[126] Y. Meyer. *Wavelets and operators*, volume 1. Cambridge university press, 1995.

[127] M. Motoo. On the Hoeffding's combinatrial central limit theorem. *Annals of the Institute of Statistical Mathematics*, 8(1):145–154, 1956.

[128] G. E. Noether. On a theorem by wald and wolfowitz. *The Annals of Mathematical Statistics*, pages 455–458, 1949.

[129] Y. Ogata. On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31, 1981.

[130] Y. Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402, 1998.

[131] K. Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175, 1900.

[132] K. Pearson. On the probability that two independent distributions of frequency are really samples from the same population. *Biometrika*, 8:250–254, 1911.

[133] D. H. Perkel, G. L. Gerstein, and G. P. Moore. Neuronal spike trains and stochastic point processes. ii Simultaneous spike trains. *Biophysical Journal*, 7:419–440, 1967.

[134] V. Pernice, B. Staude, S. Cardanobile, and S. Rotter. How structure determines correlations in neuronal networks. *PLoS Computational Biology*, 7(5):e1002059, 2012.

[135] F. Pesarin and L. Salmaso. *Permutation tests for complex data: Theory, Applications and Software*. Series in Probability and Statistics. Wiley, 2010.

[136] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. J. Chichilnisky, and E. P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

[137] G. Pipa, M. Diesmann, and S. Grün. Significance of joint-spike events based on trial-shuffling by efficient combinatorial methods. *Complexity*, 8(4):79–86, 2003.

[138] G. Pipa and S. Grün. Non-parametric significance estimation of joint-spike events by shuffling and resampling. *Neurocomputing*, 52:31–37, 2003.

[139] E. J. Pitman. Significance tests which may be applied to samples from any populations. *Journal of the Royal Statistical Society*, 4:117–130, 1937.

[140] E. J. Pitman. Significance tests which may be applied to samples from any populations. II. The correlation coefficient test. *Journal of the Royal Statistical Society*, 4:225–232, 1937.

[141] E. J. Pitman. Significance tests which may be applied to samples from any populations. III. The analysis of variance test. *Biometrika*, 29:322–335, 1938.

[142] C. Pouzat and A. Chaffiol. Automatic spike train analysis and report generation. An implementation with R, R2HTML and STAR. *Journal of neuroscience methods*, 181(1):119–144, 2009.

[143] J. Præstgaard and J. A. Wellner. Exchangeably weighted bootstraps of the general empirical process. *The Annals of Probability*, pages 2053–2086, 1993.

[144] M. H. Quenouille. Approximate tests of correlation in time-series. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 45, pages 483–484. Cambridge Univ Press, 1949.

[145] P. Reynaud-Bouret. Adaptive estimation of the intensity of inhomogeneous poisson processes via concentration inequalities. *Probability Theory and Related Fields*, 126(1):103–153, 2003.

[146] P. Reynaud-Bouret and V. Rivoirard. Near optimal thresholding estimation of a poisson intensity on the real line. *Electronic journal of statistics*, 4:172–238, 2010.

[147] P. Reynaud-Bouret, V. Rivoirard, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *Journal of Mathematical Neuroscience*, 4(3), 2014.

[148] P. Reynaud-Bouret, V. Rivoirard, and C. Tuleau-Malot. Inference of functional connectivity in neurosciences via Hawkes processes. In *1st IEEE Global Conference on Signal and Information Processing*, 2013, Austin Texas.

[149] P. Reynaud-Bouret and S. Schbath. Adaptive estimation for hawkes processes; application to genome analysis. *The Annals of Statistics*, 38(5):2781–2822, 2010.

[150] A. Riehle, F. Grammont, M. Diesmann, and S. Grün. Dynamical changes and temporal precision of synchronized spiking activity in monkey motor cortex during movement preparation. *Journal of Physiology-Paris*, 94(5):569–582, 2000.

[151] A. Riehle, F. Grammont, and W. A. MacKay. Cancellation of a planned movement in monkey motor cortex. *Neuroreport*, 17(3):281–285, 2006.

[152] V. Rivoirard. Nonlinear estimation over weak besov spaces and minimax bayes method. *Bernoulli*, 12(4):609–632, 2006.

[153] J. P. Romano. Bootstrap and randomization tests of some nonparametric hypotheses. Technical Report 270, Department of Statistics - Stanford University, 1987.

[154] J. P. Romano. A bootstrap revival of some nonparametric distance tests. *Journal of the American Statistical Association*, 83(403):698–708, 1988.

[155] J. P. Romano. Bootstrap and randomization tests of some nonparametric hypotheses. *The Annals of Statistics*, 17(1):141–159, 1989.

[156] J. P. Romano and M. Wolf. Exact and approximate step-down methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108, 2005.

[157] S. M. Ross. *Introduction to probability models*. Academic press, 9th edition, 2006.

[158] H. Rubin and R. A. Vitale. Asymptotic distribution of symmetric statistics. *Ann. Statist.*, pages 165–170, 1980.

[159] L. Sansonnet. *Inférence non-paramétrique pour des interactions poissoniennes*. PhD thesis, Université Paris Sud, 2013.

[160] L. Sansonnet and C. Tuleau-Malot. A model of poissonian interactions and detection of dependence. *Statistics and Computing*, 25(2):449–470, 2015.

[161] H. Scheffe. Statistical inference in the non-parametric case. *The Annals of Mathematical Statistics*, 14(4):305–332, 1943.

[162] R. L. Schilling. *Measures, integrals and martingales*. Cambridge University Press, 2005.

[163] W. Schneller. A short proof of motoo's combinatorial central limit theorem using stein's method. *Probability theory and related fields*, 78(2):249–252, 1988.

[164] R. J. Serfling. *Approximation theorems of mathematical statistics*. John Wiley & Sons, 2009.

[165] C. P. Shapiro and L. Hubert. Asymptotic normality of permutation statistics derived from weighted sums of bivariate functions. *The Annals of Statistics*, pages 788–794, 1979.

[166] W. Singer. Synchronization of cortical activity and its putative role in information processing and learning. *Annual Review of Physiology*, 55(1):349–374, 1993.

[167] V. G. Spokoiny. Adaptive hypothesis testing using wavelets. *The Annals of Statistics*, 24(6):2477–2498, 1996.

[168] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques*, 81(1):73–205, 1995.

[169] W. Truccolo, U. T. Eden, Ma. R. Fellows, J. P. Donoghue, and E. N. . A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of neurophysiology*, 93(2):1074–1089, 2005.

[170] C. Tuleau-Malot, A. Rouis, F. Grammont, and P. Reynaud-Bouret. Multiple Tests based on a Gaussian Approximation of the Unitary Events method with delayed coincidence count. *Neural computation*, 26(7):1408–1454, 2014.

[171] P. Uhlhaas, G. Pipa, B. Lima, L. Melloni, S. Neuenschwander, D. Nikolić, and W. Singer. Neural synchrony in cortical networks: history, concept and current status. *Frontiers in integrative neuroscience*, 3:17, 2009.

[172] A. W. van der Vaart. *Asymptotic statistics*. Cambridge series in Statistical and Probabilistic Mathematics, 1998.

[173] A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer, New York, 1996.

[174] V. S. Varadarajan. On the convergence of sample probability distributions. *Sankhyà*, 19(1–2):23–26, 1958.

[175] V. Ventura. Bootstrap tests of hypotheses. In *Analysis of parallel spike trains*, pages 191–220. Springer Series in Computational Neuroscience, 2010.

[176] V. Ventura, C. Cai, and R. E. Kass. Statistical assessment of time-varying dependency between two neurons. *Journal of Neurophysiology*, 94(4):2940–2947, 2005.

[177] D. Vere-Jones and T. Ozaki. Some examples of statistical estimation applied to earthquake data. *Annals of the Institute of Statistical Mathematics*, 34(1):189–207, 1982.

[178] A. Wald and J. Wolfowitz. Statistical tests based on permutations of the observations. *The Annals of Mathematical Statistics*, 15(4):358–372, 1944.

[179] J. Wolfowitz. Additive partition functions and a class of statistical hypotheses. *The Annals of Mathematical Statistics*, 13(3):247–279, 1942.

[180] A. F. Yodé. Asymptotically minimax test of independence. *Mathematical Methods of Statistics*, 13(2):201–234, 2004.

[181] A. F. Yodé. Adaptive minimax test of independence. *Mathematical Methods of Statistics*, 20(3):246–268, 2011.

Développement (français)

Motivation biologique. Cette thèse est motivée par la problématique de détection de synchronisations en neurosciences. Plus précisément, les neurones communiquent entre eux grâce aux potentiels d'actions qui sont des dépolarisations brèves et stéréotypées du potentiel de membrane. Nous appelons la suite des temps d'occurrence des potentiels d'action un *train de spikes* (les *spikes* étant les temps d'occurrence eux mêmes). Même si, en raison du degré de résolution, les enregistrements des trains de spikes sont discrétisés dans le temps, leur dimension est si grande qu'il n'est ni réaliste, ni raisonnable de les modéliser par des variables ou des vecteurs aléatoires réels. Il semble plus naturel de les modéliser par des processus ponctuels.

Récemment, les neurobiologistes ont découvert que les neurones communiquent entre eux, non seulement en émettant un grand nombre de potentiels d'action par unité de temps, mais également en coordonnant leur activité afin de transmettre l'information plus rapidement et en utilisant moins d'énergie. Il est communément admis que de tels phénomènes, appelés *synchronisations*, jouent un rôle important dans l'activité cérébrale. Une première étape vers une meilleure compréhension de la transmission du message nerveux est donc de pouvoir détecter les moments où apparaissent de telles synchronisations, et ainsi, déterminer les éventuelles connectivités fonctionnelles (interactions entre les neurones). De nos jours, il est possible, d'un point de vue biologique, d'avoir accès aux trains de spikes de plusieurs neurones enregistrés simultanément. Par ailleurs, de nombreuses méthodes statistiques (voir par exemple la méthode *Unitary Events* (UE) de Grün [67] et toutes ses dérivées) ont été développées dans la littérature en neurosciences. Elles sont généralement basées sur différentes notions de coïncidence, apparaissant intuitivement à chaque fois que les neurones étudiés déchargent un potentiel d'action quasiment en même temps. De plus, elles sont généralement décomposées en deux étapes : la première consiste à construire un test d'indépendance sur une petite plage de temps donnée, et la seconde consiste à l'appliquer simultanément sur des fenêtres glissantes recouvrant l'intervalle de temps de l'expérience. Les fenêtres sur lesquelles l'indépendance est rejetée désignent les moments où il y a significativement trop (ou pas assez) de coïncidences. Différentes approches statistiques ont été considérées. La première s'appuie sur des modèles pour les processus. Par exemple, la méthode *Multiple Tests based on a Gaussian Approximation of the Unitary Events* (MTGAUE), récemment introduite par Tuleau-Malot et ses co-auteurs [170], est basée sur l'hypothèse que les trains de spikes sont des processus de Poisson homogènes. Cependant, aucun modèle n'est communément accepté à ce jour par les neuroscientifiques, et les hypothèses de stationnarité ne sont clairement pas vérifiées par de nombreux jeux de données expérimentales. La seconde approche considère plutôt des méthodes par rééchantillonnage, évitant ainsi des hypothèses de modèles. Par exemple, basée sur une approche de type bootstrap, la méthode *Trial-Shuffling* (TS) introduite par Pipa et ses co-auteurs [137, 138], détruit les structures de dépendance éventuelles en tirant des couples de processus ponctuels issus d'essais différents. Cependant, nous illustrons dans une étude par

simulation, que si la statistique bootstrappée n'est pas centrée, cette approche peut mener à des tests trop conservatifs, comme c'est le cas pour le TS. Dans la lignée de ces travaux, nous souhaitons donc construire des tests d'indépendance non-paramétriques basés sur des approches de type bootstrap ou permutation, applicables aux processus ponctuels, qui soient validés à la fois théoriquement et en pratique. Avant d'introduire de telles procédures et leurs études, précisons que pour la seconde étape (celle de test multiples), nous avons adapté la célèbre méthode de Benjamini et Hochberg [16] à notre cas. Puis, avant de l'appliquer à des données expérimentales, nous avons vérifié par simulation que la méthode de test multiple est performante d'un point de vue pratique.

Construction des tests d'indépendance, et étude asymptotique. Étant donnée l'observation d'un échantillon $\mathbb{X}_n = (X_1, \dots, X_n)$ de n copies indépendantes et identiquement distribuées $X_i = (X_i^1, X_i^2)$ (pour $1 \leq i \leq n$) d'une variable aléatoire $X = (X^1, X^2)$ de loi inconnue P, nous souhaitons tester l'hypothèse nulle $(\mathcal{H}_0)$: "les coordonnées X^1 et X^2 de X sont indépendantes" contre l'alternative $(\mathcal{H}_1)$: "elles ne le sont pas", c'est-à-dire

$$(\mathcal{H}_0)\ P = P^1 \otimes P^2 \quad \text{contre} \quad (\mathcal{H}_1)\ P \neq P^1 \otimes P^1,$$

où P^1 et P^2 désignent les marginales de la loi P.
Les tests d'indépendance non-paramétriques ont été largement étudiés dans la littérature statistique. Nous nous sommes inspirés des approches par bootstrap et permutation de Romano [155] (lui-même inspiré de Hoeffding [84] pour la permutation) et de van der Vaart et Wellner [173]. Notre statistique de test est une U-statistique renormalisée, supposée non-dégénérée, de noyau h, définie par

$$\sqrt{n} U_{n,h}(\mathbb{X}_n) = \frac{\sqrt{n}}{n(n-1)} \sum_{i \neq j} h(X_i, X_j).$$

En raison de la nécessité du recentrage pour les approches par bootstrap, nous supposons que sous l'hypothèse nulle, cette statistique de test est d'espérance égale à zéro. Un cas particulier important dans cette thèse, appelé *cas linéaire*, consiste à prendre le noyau de la forme h_φ défini pour tout $x = (x^1, x^2)$ et $y = (y^1, y^2)$ par

$$h_\varphi(x, y) = \frac{1}{2} \left[\varphi(x^1, x^2) + \varphi(y^1, y^2) - \varphi(x^1, y^2) - \varphi(y^1, x^2) \right],$$

où φ est une fonction mesurable à valeurs réelles. Dans ce cas, la U-statistique est un estimateur sans biais de $\int \varphi(x^1, x^2)[dP(x^1, x^2) - dP^1(x^1) dP^2(x^2)]$, et l'hypothèse de recentrage est automatiquement vérifiée. Nous pouvons remarquer que notre statistique de test dans le *cas linéaire* ressemble à celle de type Kolmogorov-Smirnov de Romano et de van der Vaart et Wellner, mais sans un supremum. Une des nouveautés de cette thèse est que nous ne supposons pas nécessairement que φ est de la forme produit, comme c'est souvent le cas dans la littérature, afin de pouvoir l'appliquer à des fonctionnelles adaptées aux neurosciences telles que la notion de coïncidence.
Il reste donc à calibrer les valeurs critiques définissant la zone de rejet. Fixons un niveau prescrit α dans $]0, 1[$. Par soucis de simplicité, nous présentons uniquement les tests unilatéraux à droite, mais ceux unilatéraux à gauche et bilatéraux sont également étudiés dans cette thèse. Un premier résultat permettant de construire un test purement asymptotique est le théorème de la limite centrale pour les U-statistiques non-dégénérées selon lequel la loi de la statistique de test sous l'hypothèse nulle converge en distance de Wasserstein d'ordre 2 (notée d_2) vers

une loi normale centrée de variance $\sigma^2_{P^1\otimes P^2}$ dépendant uniquement des marginales P^1 et P^2, ou plus précisément, si $P = P^1 \otimes P^2$, alors

$$d_2\left(\mathcal{L}\left(\sqrt{n}U_{n,h}, P^1\otimes P^2 \middle| \mathbb{X}_n\right), \mathcal{N}\left(0, \sigma^2_{P^1\otimes P^2}\right)\right) \underset{n\to+\infty}{\longrightarrow} 0.$$

Une idée naturelle est de diviser la statistique de test par un estimateur de la variance, et de rejeter lorsque ce quotient est supérieur au quantile d'ordre $(1-\alpha)$ de la loi normale centrée réduite. Cependant, généralement en biologie, peu de données sont exploitables, et une procédure purement asymptotique ne semble pas adaptée.

Nous nous sommes donc tournés vers des méthodes basées sur du rééchantillonnage, telles que le bootstrap ou la permutation. L'idée générale est de construire un nouvel échantillon $\tilde{\mathbb{X}}_n$ à partir de l'observation $\mathbb{X}_n$, de loi notée ici $\tilde{P}_n$ (dépendant de $\mathbb{X}_n$), de sorte que la loi conditionnelle de la statistique de test rééchantillonnée $\sqrt{n}U_{n,h}(\tilde{\mathbb{X}}_n)$ sachant $\mathbb{X}_n$, notée $\mathcal{L}(\sqrt{n}U_{n,h}, \tilde{P}_n|\mathbb{X}_n)$ approche (généralement asymptotiquement) la loi de la statistique de test sous l'hypothèse nulle. Si ceci reste vrai lorsque $\mathbb{X}_n$ ne vérifie pas $(\mathcal{H}_0)$, on dit alors que la méthode de rééchantillonnage permet de *reconstruire la loi sous* $(\mathcal{H}_0)$. L'approche par bootstrap considère l'échantillon dit bootstrappé, noté $\mathbb{X}^*_n$, de loi le produit des marginales empiriques, à savoir $P^1_n \otimes P^2_n$ où $P^j_n = n^{-1}\sum_{i=1}^n \delta_{X^j_i}$ pour $j = 1, 2$. Cela consiste simplement à tirer avec remise chaque coordonnés indépendamment. L'approche par permutation consiste à permuter uniformément les secondes coordonnés, à savoir, $\mathbb{X}^{\Pi_n}_n = (X^{\Pi_n}_1, \ldots, X^{\Pi_n}_n)$ avec pour tout i dans $\{1, \ldots, n\}$, $\mathbb{X}^{\Pi_n}_i = (X^1_i, X^2_{\Pi_n(i)})$ où Π_n est une permutation uniforme de $\{1, \ldots, n\}$. Nous notons sa loi $P^\star_n$. Pour chacune des approches ($\tilde{P} = P^1_n \otimes P^2_n$ ou $\tilde{P} = P^\star_n$), nous obtenons que

$$d_2\left(\mathcal{L}\left(\sqrt{n}U_{n,h}, \tilde{P}_n \middle| \mathbb{X}_n\right), \mathcal{N}\left(0, \sigma^2_{P^1\otimes P^2}\right)\right) \underset{n\to+\infty}{\longrightarrow} 0,$$

et ce, que $\mathbb{X}_n$ vérifie ou non $(\mathcal{H}_0)$. La loi limite étant celle de la statistique sous hypothèse nulle, nous avons bien démontré que les deux approches permettent de reconstruire la loi sous $(\mathcal{H}_0)$. Remarquons que pour la permutation, nous avons dû nous restreindre au *cas linéaire* pour des raisons techniques. Cependant, même dans ce cas particulier, ce résultat est l'un des plus novateurs de cette thèse, car il reste vrai sous l'alternative où les variables permutées ne sont plus échangeables. En ce sens, il répond partiellement (nous ne considérons pas de supremum) à une question ouverte laissée par van der Vaart et Wellner dans [173].

Ceci nous permet aisément d'en déduire que les tests par bootstrap et par permutation prenant comme valeurs critiques les quantiles conditionnels d'ordre $(1-\alpha)$ de $\mathcal{L}(\sqrt{n}U_{n,h}, \tilde{P}_n|\mathbb{X}_n)$ pour $\tilde{P} = P^1_n \otimes P^2_n$ et $\tilde{P} = P^\star_n$ ont de bonnes propriétés asymptotiques, à savoir qu'ils sont *asymptotiquement de taille prescrite* α, i.e. la probabilité sous $(\mathcal{H}_0)$ de rejeter l'indépendance à tort converge vers α lorsque n tend vers $+\infty$, et qu'ils sont *puissants contre les alternatives* sous lesquelles l'espérance de la statistique de test (nulle sous $(\mathcal{H}_0)$) est strictement positive, i.e. sous ces alternatives, la probabilité de rejeter en ayant raison converge vers 1 lorsque n tend vers $+\infty$.

Les calculs des quantiles conditionnels étant généralement très coûteux, nous les approchons en pratique par des méthodes de Monte Carlo. Nous avons démontré que les tests obtenus en remplaçant les quantiles exacts par ceux approchés vérifient les mêmes propriétés asymptotiques, à savoir qu'ils sont asymptotiquement de taille prescrite, et puissants contre les mêmes alternatives à condition que le nombre d'étapes dans l'approche par Monté Carlo tende vers $+\infty$ avec n.

Une étude par simulation nous a permis de vérifier la validité de nos tests d'un point de vue pratique, ainsi que de les comparer à des méthodes classiques en neurosciences telles que les méthodes MTGAUE ou TS.

Cependant, comme mentionné ci-dessus, même si nos tests par bootstrap ou par permutation semblent être plus performants d'un point de vue pratique que le test "naïf" basé simplement sur le théorème de la limite centrale pour les U-statistiques, le nombre d'observations en biologie est généralement trop petit pour que des résultats purement asymptotiques permettent de justifier pleinement d'un point de vue théorique l'application de telles procédures. Nous nous sommes donc également intéressés à l'étude de leur propriétés non-asymptotiques.

Étude non-asymptotique et inégalité de concentration. Les tests par permutation sont construits de manière à être exactement de niveau non-asymptotique prescrit, c'est-à-dire que, quelque soit le nombre d'observations, la probabilité sous $(\mathcal{H}_0)$ de rejeter à tort est toujours contrôlée par α. En ce sens, il est communément admis que lorsque les deux approches sont possibles, il vaut mieux privilégier celle par permutation. Pour cela, nous n'étudions que la permutation d'un point de vue non-asymptotique. Nous nous restreignons au *cas linéaire* dans lequel la statistique de test permutée s'écrit comme une somme permutée recentrée, ou plus précisément

$$\sqrt{n}U_{n,h_\varphi}\left(\mathbb{X}_n^{\Pi_n}\right) = \frac{\sqrt{n}}{n-1}\left(\sum_{i=1}^n \varphi\Big(X_i^1, X_{\Pi_n(i)}^2\Big) - \mathbb{E}\left[\sum_{i=1}^n \varphi\Big(X_i^1, X_{\Pi_n(i)}^2\Big)\middle|\mathbb{X}_n\right]\right).$$

Pour nous ramener à un contexte plus classique, l'étude non-asymptotique est développée dans un cadre de variables aléatoires à densité dans $\mathbb{L}_2([0,1]^2)$. Dans un but d'adaptativité, nous avons introduit une procédure de tests agrégés basée sur une méthode de seuillage par ondelettes. L'idée est de décomposer la différence entre la densité jointe et le produit de ses marginales dans une base d'ondelettes (ici, la base de Haar), et de construire pour chaque coefficient, un test basé sur la statistique de test introduite précédemment dans le *cas linéaire* où la fonction φ n'est autre que l'ondelette correspondante, et sur la même approche par permutation. La procédure de tests agrégés applique simultanément chaque test (à un niveau corrigé pour tenir compte de la multiplicité des tests) jusqu'à une certaine échelle (dépendant de n), et rejette l'indépendance si au moins un des tests individuels la rejette. Nous étudions les vitesses de séparation uniformes d'un point de vue minimax (cf [13]).
Pour cela, nous avons d'abord démontré une inégalité de concentration de type Bernstein pour les sommes permutées aléatoirement. Les inégalités de concentration sont des outils très fins permettant de contrôler la probabilité qu'une fonction d'une où plusieurs variables aléatoires s'écarte de sa médiane ou de sa moyenne. Il existe de nombreuses inégalités de concentrations pour des fonctions d'une permutation aléatoire (voir par exemple [124, 2, 33]). Nous avons suivi l'idée de Adamczak et de ses co-auteurs [2], fondée sur l'inégalité fondamentale de Talagrand pour les sommes permutées [168, Théorème 5.1]. Cela nous a mené à une première inégalité de concentration pour la racine de la somme permutée. Celle-ci n'étant pas suffisamment fine, nous l'avons réutilisée dans le déroulement de la preuve afin d'améliorer le résultat obtenu. Cela nous a fourni une inégalité de concentration exponentielle autour de la médiane. Ledoux [114] a montré que dans ce cas, la moyenne est équivalente à la médiane et nous avons alors obtenu l'inégalité suivante : pour tous réels $a_1, \dots, a_n$ et toute permutation aléatoire de $\{1, \dots, n\}$ uniforme, notée Π_n,

$$\mathbb{P}\left(\left|\sum_{i=1}^n a_{i,\Pi_n(i)} - \mathbb{E}\left[\sum_{i=1}^n a_{i,\Pi_n(i)}\right]\right| \geq 2\sqrt{2\left(\frac{1}{n}\sum_{i,j=1}^n a_{i,j}^2\right)x} + 2\max_{1\leq i,j\leq n}|a_{i,j}|\,x\right) \leq c_0 e^{-c_1 x},$$

où c_0 et c_1 sont des constantes universelles strictement positives.

Ce résultat de concentration nous a alors permis d'obtenir un contrôle suffisamment fin des quantiles conditionnels pour majorer à constante près la vitesse de séparation uniforme sur des Besov faibles de paramètre γ, en distance $\mathbb{L}_2$ de la procédure de tests agrégés par $(n/\ln(n))^{\frac{-\gamma}{2\gamma+2}}$. Même si les bornes infimum ne sont pas connues à ce jour, au vue de la littérature, et notamment des vitesses obtenues par Ingster [94] en distances $\mathbb{L}_p$ sur des espaces de régularité de type Hölder, la vitesse de séparation uniforme obtenue semble être optimale d'un point de vue minimax et donc adaptative en ce sens (puisque la procédure ne dépend pas du paramètre de régularité du Besov faible γ).